NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

A COMBINED REGRESSION AND BOX-JENKINS
PREDICTION MODEL FOR REENLISTMENT
IN SELECTED NAVY RATINGS

by

Kevin J. Sherry

June 1984

Thesis Advisors: George W. Thomas
William E. McGarvey

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Combined Regression and Box-Jenkins Prediction Model for Reenlistment In Selected Navy Ratings		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Kevin Joseph Sherry		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943		12. REPORT DATE June 1984
		13. NUMBER OF PAGES 79
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Time Series; Box-Jenkins methodology; Univariate analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This thesis examines five Navy ratings using two distinct modeling techniques in an effort to predict first term reenlistments. The techniques utilized are Box-Jenkins time series analysis and linear regression. A combined model utilizing both techniques is also developed. The ability of the models to predict is considered adequate for three of the five ratings and not adequate for two of the ratings.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The regression models utilizing 20-24 year old unemployment as the only independent variable yielded surprisingly good predictions, once the time series patterns in the data were modeled.

Approved for public release; distribution unlimited.

A Combined Regression and Box-Jenkins Prediction Model for Reenlistment In Selected Navy Ratings

by

Kevin J. Sherry
Lieutenant Commander United States Navy
B.S., Fordham University, 1971

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
June 1984

ABSTRACT

This thesis examines five Navy ratings using two distinct modeling techniques in an effort to predict first term reenlistments. The techniques utilized are Box-Jenkins time series analysis and linear regression. A combined model utilizing both techniques is also developed. The ability of the models to predict is considered adequate for three of the five ratings and not adequate for two of the ratings. The regression models utilizing 20-24 year old unemployment as the only independent variable yielded surprisingly good predictions, once the time series patterns in the data were modeled.

TABLE OF CONTENTS

I. INTRODUCTION AND REVIEW OF RELEVANT LITERATURE . . 11
A. INTRODUCTION 11
B. REVIEW OF RELEVANT LITERATURE IN THE FIELD 13
1. Overview 13
2. The Annualized Cost of Leaving (ACOL) Model 13
3. Darling's Model of Marine Corp Enlistments 14
4. Bepko's Regression Model 17
5. The Thomas-Liao Model of Careerists . . . 17

II. APPLICATION OF BOX-JENKINS ANALYSIS TO THE UNIVARIATE MODELS 18
A. INITIAL ANALYSIS OF THE DATA SETS 18
B. EXAMINATION OF AUTOCORRELATIONS AND PARTIAL AUTOCORRELATIONS 25
1. Yeoman 25
2. Storekeeper 25
3. Operations Specialist 26
4. Electronics Technicians 28
5. Boiler Technicians 30
C. MODEL DEVELOPMENT 30
1. Yeoman Model 34
2. Storekeeper Model 35
3. Operations Specialist Model 35
4. Electronics Technician Model 36
5. Boiler Technician Model 37
D. SUMMARY 38

III. THE LEADING INDICATOR, REGRESSION AND COMBINED MODELS 40
A. OVERVIEW 40
B. THE UNEMPLOYMENT MODEL 40
1. Methodology 41
C. APPLICATION TO THE SELECTED RATINGS DATA SETS 45
1. Yeoman 45
2. Storekeepers 46
3. Operations Specialists 47
4. Electronics Technicians 48
5. Boiler Technicians 49
D. LINEAR REGRESSION 50
E. MODEL VERIFICATION 50
F. THE COMBINED MODEL 53

IV. SUMMARY AND CONCLUSIONS 56
A. SUMMARY 56
B. CONCLUSIONS 58
C. SUGGESTIONS FOR FUTURE RESEARCH 59

APPENDIX A: SUMMARY OF DATA USED IN ANALYSIS 60

APPENDIX B: THE BOX-JENKINS METHOD 61
A. OVERVIEW 61
B. WHAT IS A TIME SERIES 62
C. STATIONARITY 62
D. TIME SERIES PLOT 64
E. DATA TRANSFORMATION 64
F. AUTOCORRELATION AND PARTIAL AUTOCORRELATION 67
G. THE AUTOREGRESSIVE MODEL 67
H. THE MOVING AVERAGE MODEL 68
I. THE MIXED AUTOREGRESSIVE MOVING AVERAGE MODEL 68

J. EVALUATING THE MODEL 69
1. ACF of The Residuals 70
2. Minimum Sum of Squares 70
3. t-Ratio 70
K. PARSIMONY 71
L. TRANSFER FUNCTION MODELS 71

APPENDIX C: SUMMARY OF BOX-JENKINS MODELS USED FOR ANALYSIS 73
A. MODELS USED IN UNIVARIATE ANALYSIS 73
1. Yeoman 73
2. Storekeepers 73
3. Operations Specialists 74
4. Electronics Technicians 74
5. Boiler Technicians 75
B. BOX-JENKINS MODELS OF REGRESSION RESIDUALS 75
1. Yeoman 75
2. Storekeepers 76
3. Operations Specialists 76
4. Electronics Technicians 77
5. Boiler Technicians 77

LIST OF REFERENCES 78

INITIAL DISTRIBUTION LIST 79

LIST OF TABLES

I. SUMMARY STATISTICS OF REENLISTMENT RATES 19
II. RESULTS OF UNIVARIATE MODEL FORECASTS 33
III. SUMMARY OF UNEMPLOYMENT MA 1 MODEL 44
IV. SUMMARY OF R-SQUARED AND DURBIN-WATSON STATISTICS 52
V. REGRESSION MODELS FOR SELECTED RATINGS 53
VI. FORECAST COMPARISONS OF ALL THREE MODELS 55
VII. DATA UTILIZED FOR ANALYSIS 60

LIST OF FIGURES

2.1 TIME SERIES PLOT FOR YN RATING 20
2.2 TIME SERIES PLOT FOR SK RATING 21
2.3 TIME SERIES PLOT FOR OS RATING 22
2.4 TIME SERIES PLOT FOR ET RATING 23
2.5 TIME SERIES PLOT FOR BT RATING 24
2.6 AUTOCORRELATIONS FOR YN RATING 26
2.7 PARTIAL AUTOCORRELATIONS FOR YN RATING 26
2.8 AUTOCORRELATIONS FOR SK RATING 27
2.9 PARTIAL AUTOCORRELATIONS FOR SK RATING 27
2.10 AUTOCORRELATIONS FOR OS RATING 28
2.11 PARTIAL AUTOCORRELATIONS FOR OS RATING 28
2.12 AUTOCORRELATIONS FOR ET RATING 29
2.13 PARTIAL AUTOCORRELATIONS FOR ET RATING 30
2.14 AUTOCORRELATIONS FOR BT RATING 31
2.15 PARTIAL AUTOCORRELATIONS FOR BT RATING 31
2.16 PLOT OF UNIVARIATE MODEL FOR YN RATING 34
2.17 PLOT OF UNIVARIATE MODEL FOR SK RATING 36
2.18 PLOT OF UNIVARIATE MODEL FOR OS RATING 37
2.19 PLOT OF UNIVARIATE MODEL FOR ET RATING 38
2.20 PLOT OF UNIVARIATE MODEL FOR BT RATING 39
3.1 TIME SERIES PLOT FOR UNEMPLOYMENT DATA 42
3.2 AUTOCORRELATION FUNCTION FOR UNEMPLOYMENT DATA 43
3.3 PARTIAL AUTOCORRELATIONS FOR UNEMPLOYMENT DATA 43
3.4 AUTOCORRELATION OF RESIDUALS FOR UNEMPLOYMENT MODEL 44
3.5 CROSS CORRELATION FOR YN RATING 45

3.6 CROSS CORRELATION FOR SK RATING 46
3.7 CROSS CORRELATION FOR OS RATING 47
3.8 CROSS CORRELATION FOR ET RATING 48
3.9 CROSS CORRELATION FOR BT RATING 49
B.1 THE BOX-JENKINS METHODOLOGY 63
B.2 UNITED STATES GNP 1947 - 1970 65
B.3 SQUARE ROOT TRANSFORMATION OF U.S. GNP 1947 - 1970 66
B.4 TYPICAL FORM OF AR1 MODEL ACF AND PACF 69
B.5 TYPICAL FORM OF MA1 MODEL ACF AND PACF 70
B.6 TYPICAL FORM OF ARMA 1,1 MODEL ACF AND PACF . . 72

I. INTRODUCTION AND REVIEW OF RELEVANT LITERATURE

A. INTRODUCTION

This thesis is an investigation of methods for predicting the rate of reenlistment in the armed forces, specifically the Navy. Since the advent of the all volunteer force in 1973, one of the major concerns of the military has been the retention of qualified personnel beyond their first enlistment. Referred to as first term reenlistments, this decision has been the object of extensive study and modeling by each of the services. The vast majority of this work has centered around the formulation of causal models with a heavy emphasis on economic factors. During periods when reenlistments have been below the required levels, these models have been quite good at capturing the effect of the economic factors used to suggest the level of monetary compensation. Over the past four years, this situation has changed to the point where the services are faced with such large numbers of personnel desiring to remain in the services, that high reenlistment is actually lowering the numbers of personnel that can be enlisted for some ratings at the recruiting stations. As always, there is still a need for more personnel in the nuclear related ratings but some of the less technical fields are approaching an end point where only a limited number of billets may be available to new accessions.

The object of this thesis is to attempt to construct a short term model that could aid in predicting first term reenlistments for five selected Navy ratings. Initial models will be developed utilizing the Box-Jenkins method of time series analysis. These initial models will be used to

predict the level of reenlistment for the selected ratings. Next, a leading indicator model will be developed utilizing the national unemployment rate for 20-24 year olds. Then, a refined forecast will be developed combining both time series and causal models.

The potential advantage of time series analysis lies in its simplicity and lack of reliance on external factors. To many, this is viewed as a shortcoming since it cannot explain causal relationships such as the effect of advertising dollars on sales for a company. However, the effectiveness of advertising dollars is not easy to determine and in many cases, may lead to false conclusions when used in classical regression analysis. Other errors such as the autocorrelation of an independent variable or the multicollinearity of several variables are also avoided in the use of time series analysis. Since time series analysis relies on its ability to reproduce itself over time, this allows it to be free of the errors of regression and still retain the ability to adequately forecast events: or in this case levels of reenlistment. A more thorough explanation of the Box-Jenkins method is presented in Appendix B.

If a time series model is accurate at predicting changes in trends as well as levels of reenlistment, then it may also be useful as a tool to adjust the levels of reenlistment bonuses to the most cost effective level necesarry to retain the desired force levels. If a model can accurately foresee a significant increase in reenlistments, independent of the reenlistment bonus, the bonus level for that rating can be scaled down appropriately to retain personnel without the payment of an economic rent. (Rent, in this case, is the payment of a bonus for a decision independent of the bonus award level.) To summarize, time series analysis may not prove to be a panacea in forecasting reenlistment rates but it most definitely is a tool that deserves wider consideration and application.

B. REVIEW OF RELEVANT LITERATURE IN THE FIELD

1. Overview

To the best of the author's knowledge, the application of Box-Jenkins Time Series Analysis to reenlistment models has not been previously attempted. A search, conducted by the Defense Technical Information Center (DTIC), utilizing both title and subject, did not reveal any references to its use of Box-Jenkins modeling for reenlistment. Bepko [Ref. 1], in his thesis modeling career petty officers reviews the relevant literature concerning first term reenlistment and career reenlistment from 1974 to the publication of his thesis in 1981. This review will address relevant publications since that time.

2. The Annualized Cost of Leaving (ACOL) Model

This model was developed for the Navy by John Warner at the Center for Naval Analysis and is currently the most widely used model in the Navy with relation to manpower and personnel policy decision making.

The model itself is a sophisticated multiple linear regression that attempts to capture several important underlying forces in the reenlistment decision. The general model is of the form:

$$C_{t,n} = \left[\sum_{j=t}^{n} M \,/\, (1+r)^{j-t} + \bar{W}_n + \bar{R}_n \,/\, (1+r)^{n-t} \right] - (W_t + R_t) \quad \text{(eqn 1.1)}$$

where;

1. C(t,n) = Net present value of pecuniary and non-pecuniary returns of staying in the military until time 'n' as compared to leaving at time 't'.

2. M(j) = Monetary returns to military service from period 't' through 'n'
3. W(n) = Lump sum payment of the present value of the expected post service civilian wages realized by those staying in the military until time 'n'.
4. R(n) = Lump sum payment of the present value of the expected retirement benefits realized by those staying in the military until 'n'
5. W(t) = Present value in year 't' of the expected civilian wages realized by those leaving the military in year 't'.
6. R(t) = Present value in year 't' of the expected civilian retirement payments for those leaving the military in year 't'
7. r = personal discount rate

As can be seen, this model relies on development of several sub-models relating to civilian wage structure, future policy decisions concerning lump sum bonus payments and an individual's personal discount rate over time. While the results obtained with this model have been superb, the model itself is complex and somewhat difficult. Since the introduction of the general model described above, it has been further refined to include a 'taste' factor for military service and a random disturbance term which is used to capture the effects of sea shore rotation, poor duty station and family separation. These factors improve the model but at a cost of ever increasing complexity.

3. Darling's Model of Marine Corp Enlistments

Darling [Ref. 2], utilized a combination of Box-Jenkins time series analysis and multiple linear regression to predict the supply of upper mental category recruits to the Marine Corp. The procedure entailed development of three separate models using two distinct techniques,

the first was a standard multiple regression of the logit form:

$$S(M) = a \cdot M_0 / \left[b M_0 + (a - b M_0) e^{-a(M - M_0)} \right] \quad \text{(eqn 1.2)}$$

where;

1. S = The supply of military recruits
2. M = The military wage
3. a = The stochastic error term

In this phase of development, the following variables were introduced:

1. Civilian military pay ratio
2. National unemployment for 16-19 year old males
3. Monthly leads from print media
4. Number of Marine recruiters
5. Dummy variable to account for anomalies in a particular time period.

Further analysis led to inclusion of items one, two and five from the above list in the final regression model.

Next, a totally independent model based on Box-Jenkins methodology was developed for Marine accessions. This was a univariate model whose purpose was to capture the effect of seasonality that was missed by the regression model. The Box-Jenkins method yielded a model of the form:

$$Y_t = \phi_1 Y_{t-1} - \phi_1 Y_{t-13} + Y_{t-12} + \theta_0 + e_t - \theta_1^{\,S} e_{t-12} \quad \text{(eqn 1.3)}$$

where;

1. Y = The number of high school graduates in mental categories I and II that enlist in the Marine Corp in month t.

2. ϕ = The autoregressive coefficient that describes the model
3. θ = The moving average coefficient that describes the model
4. e = The error term of the model.

These two models were initially used to generate independent forecasts of the actual enlistment supply. Both models were adequate in capturing trends in the actual enlistment rate but were somewhat deficient in the actual numbers generated. At this point, Darling combined the techniques by utilizing the Box-Jenkins method to take advantage of the high degree of serial correlation remaining in his regression models residuals by modeling the residuals as a separate time series. The result of this method was to predict 'error' terms to be applied to the results of the regression equation. This combined model was summarized by .eqn 1.4 shown below;

$$LSVc = LSVmr + Zt' \qquad \text{(eqn 1.4)}$$

where;

1. LSVc = The predicted value of the combined model
2. LSVmr = The predicted value using the regression model only
3. Zt' = The Box-Jenkins model of the residuals of the regression model

The resulting combined forecasts were substantially better than either of the techniques could achieve separately and was extremely accurate in capturing the general trend of the enlistment supply.

4. Bepko's Regression Model

Bepko [Ref. 1], concentrated his efforts in developing a multiple regression model for forecasting career retention beyond the first enlistment. This model was unique in that it evaluated the Navy Rating structure by occupational groupings and utilized an age specific index for introducing unemployment into the model. This was the first work to look at ratings and unemployment together in this manner.

Bepko's [Ref. 1], overall findings were that evaluating the reenlistment decision by groupings among careerists yielded more relevant models for the application of bonus payments and that unemployment among the 25-39 age group was a very significant factor in the reenlistment decision of careerists.

5. The Thomas-Liao Model of Careerists

The work of Thomas and Liao [Ref. 3], continued along the same track as Bepko [Ref. 1], in the examination of the reenlistment rate among careerists, or those considering their second or subsequent reenlistment. The difference in this model is the unique grouping consideration given to the rating structure, where the ratings are aggregated by the patterns of their past reenlistment percentages. The effect of this appears to better

capture the effect of the significant variables on the reenlistment decision. The variables utilized in this study were national unemployment, the civilian military pay ratio and tenure as expressed by years of service. The results of the predictions generated by the regression equation were excellent and generally less than 10% in total error.

II. APPLICATION OF BOX-JENKINS ANALYSIS TO THE UNIVARIATE MODELS

This chapter presents the analysis of re-enlistment data by the Box- Jenkins technique. Box-Jenkins analysis involves three steps;

1. Identification - This involves analysis of the time series plots of the raw data to try and discern any obvious trend or seasonality
2. Estimate - This step involves analysis of the auto-correlations and partial autocorrelations to provide an estimate for an initial model
3. Forecast - This step involves running the models and generating predictions which are then evaluated for their adequacy.

Should any model prove inadequate, the estimation of the model is re-evaluated to find a more suitable model.

A. INITIAL ANALYSIS OF THE DATA SETS

The Box-Jenkins method was applied to data sets of the number of first term reenlistees for the following ratings:

1. Yeoman (YN)
2. Storekeeper (SK)
3. Operations Specialist (OS)
4. Electronics Technician (ET)
5. Boiler Technician (BT)

These ratings were selected for analysis because they presented a representative mix in mental category groupings, varying degrees of general and specific training and also provided sufficient numbers of reenlistments to perform valid analysis.

The data consisted of monthly summaries of first term reenlistment percentages for the subject ratings covering the period from October, 1980 through September, 1983, these data were provided by Mr. James McEwan, the statistician for the Re-enlistment Programs Development Office (OP-136), a summary of the data is shown in table I. Reenlistment percentages for these ratings ranged from a low of 41.2 per month (BT), to a high of 91.2 per month (ET).[1] It should be noted here that the time series plots of the raw data neither suggest any clear cut trends or seasonality, nor are the series similar across the ratings.

The time series plots on the following pages represent the percentages of reenlistments in each of the selected ratings, the vertical axis is the reenlistment percentage and the horizontal axis is the time line. The origin for the time line is October, 1980 and the end point is September, 1983.

TABLE I

SUMMARY STATISTICS OF REENLISTMENT RATES

RATE	DATES	MEAN	ST.DEV.
YN	10/80-8/83	55.5	11.9
SK	10/80-9/83	49.5	13.2
OS	10/80-9/83	43.6	18.0
ET	10/80-9/83	91.4	06.5
BT	10/80-9/83	41.2	14.7

---

[1]These numbers are to be viewed in a relative sense for their ability to provide sufficient population size for their respective models and not as any measurement of "health" in a particular rating. This is to say that these percentages do not necessarily measure need in a particular rating, but rather the percentage of eligibles reenlisting In addition, time series plots of each data set are presented in Fig. 2.1 through Fig. 2.5

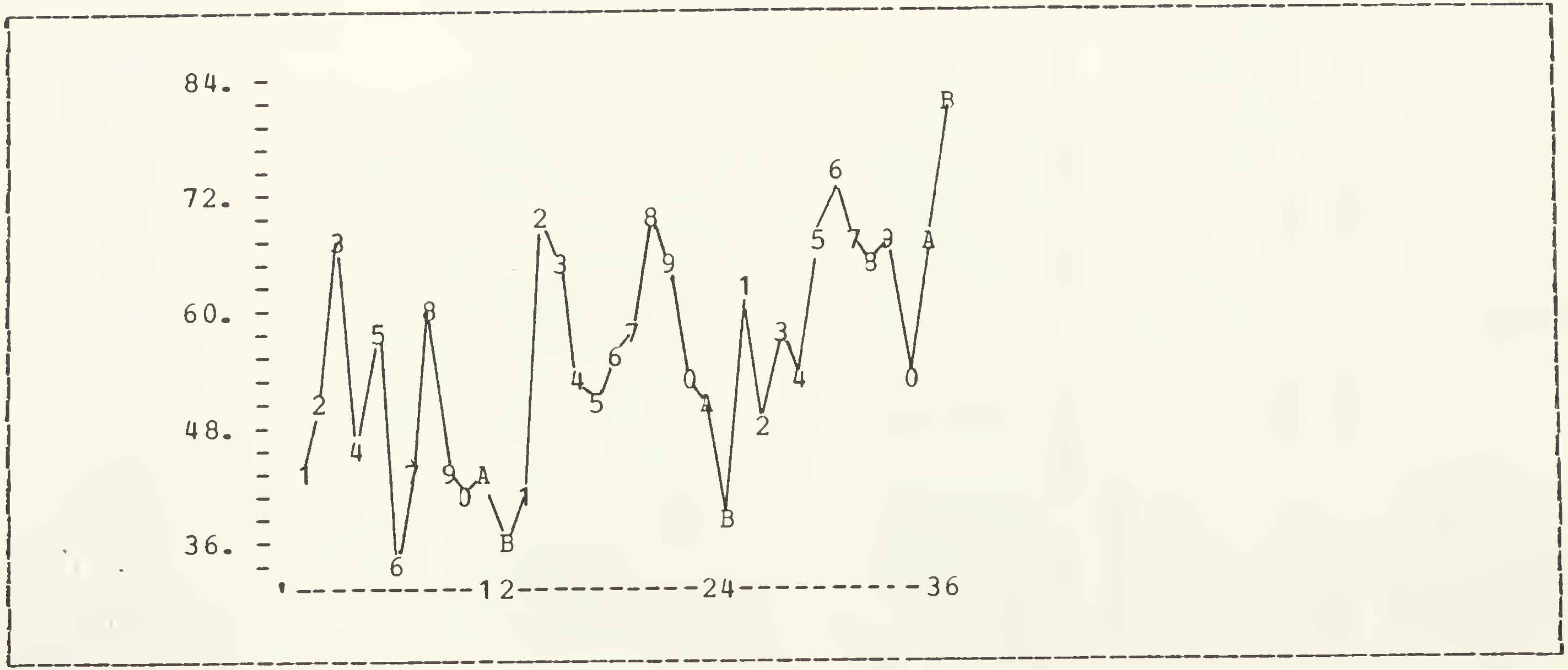

Figure 2.1 TIME SERIES PLOT FOR YN RATING.

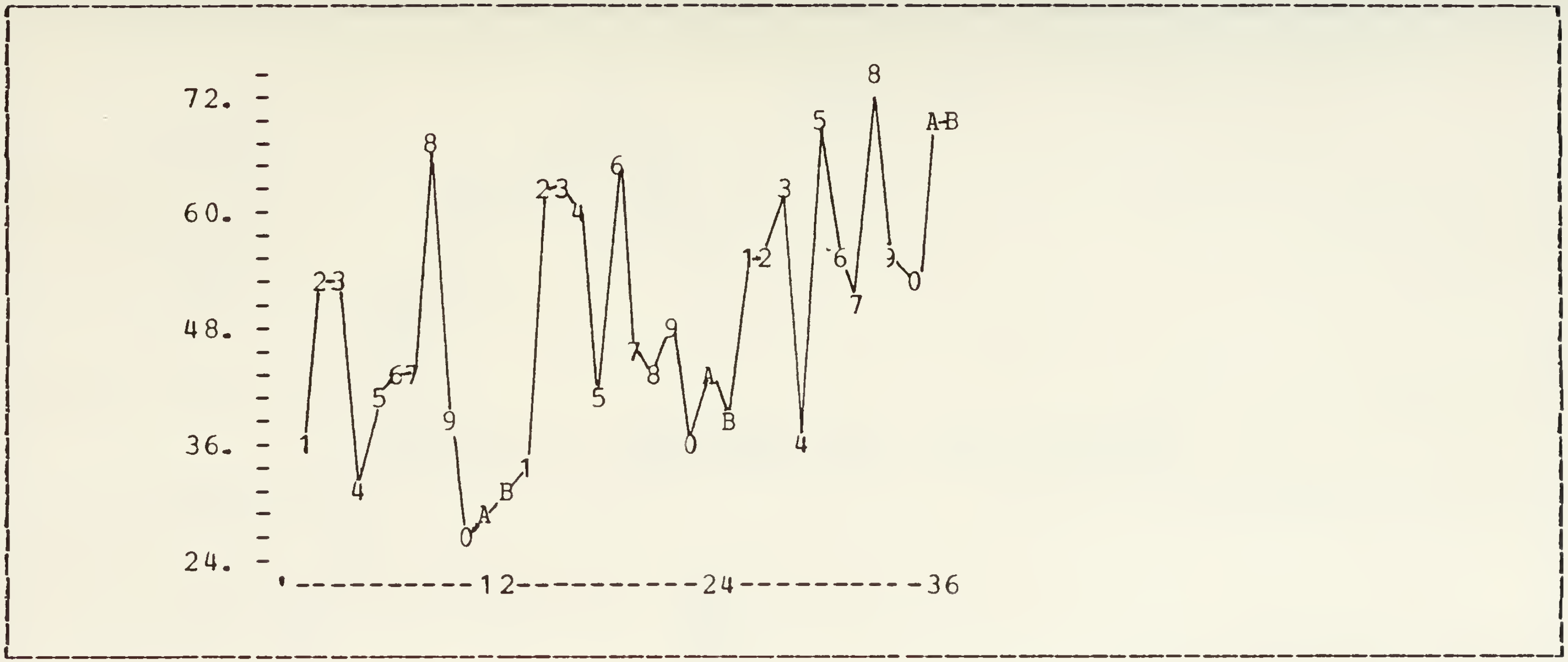

Figure 2.2 TIME SERIES PLOT FOR SK RATING.

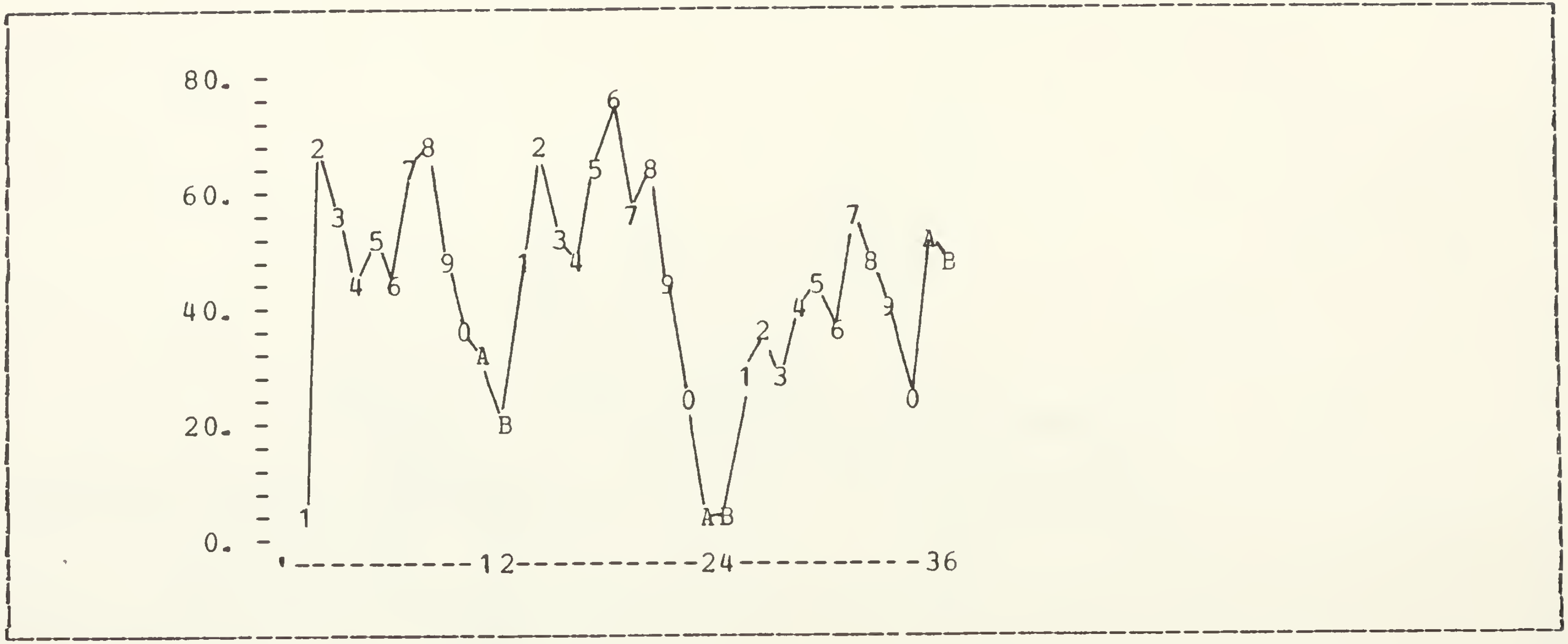

Figure 2.3 TIME SERIES PLOT FOR OS RATING.

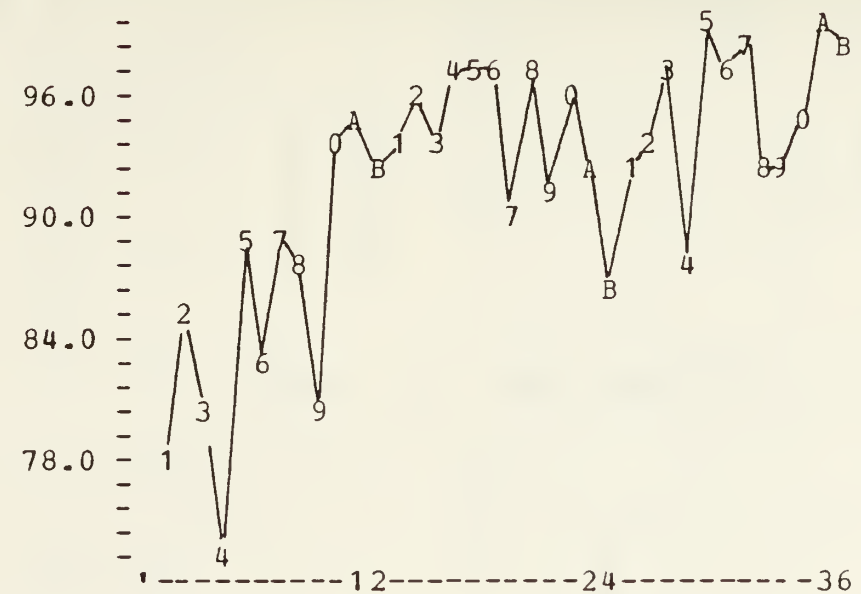

Figure 2.4 TIME SERIES PLOT FOR ET RATING.

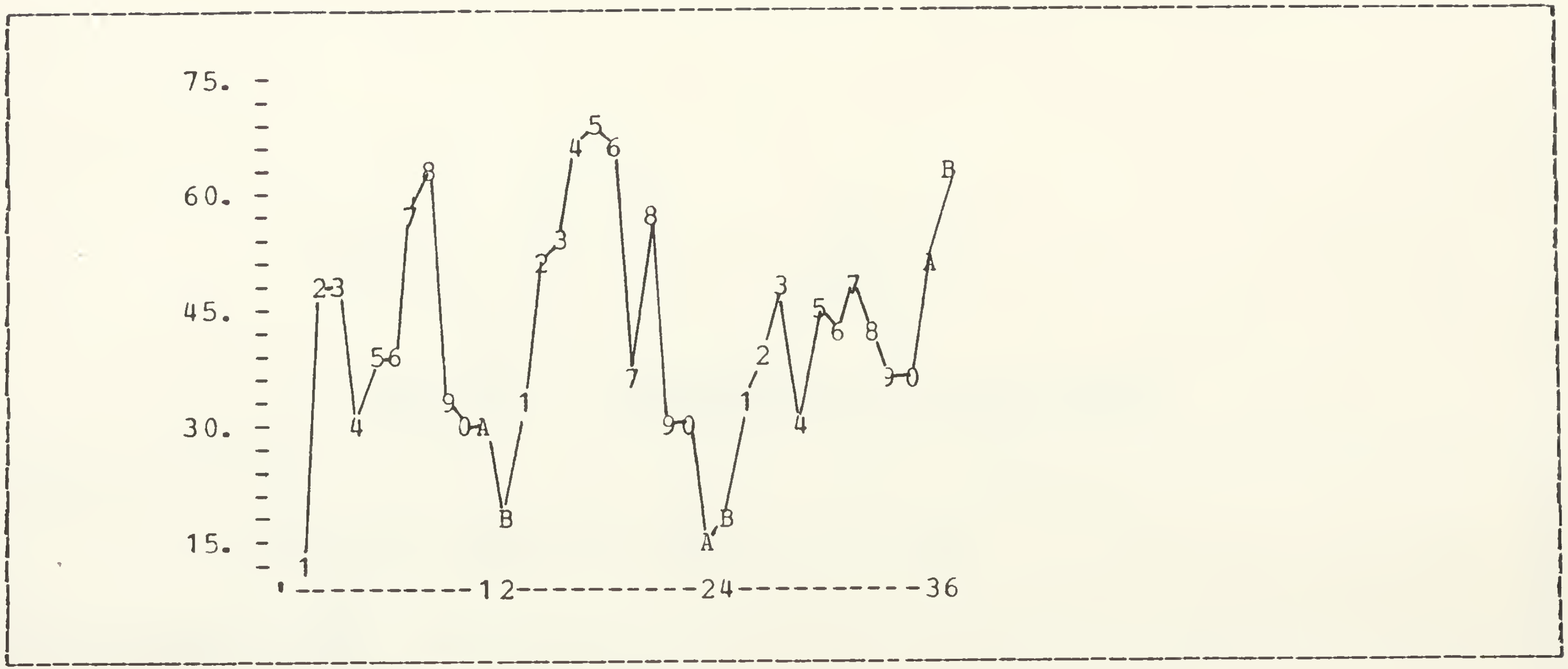

Figure 2.5   TIME SERIES PLOT FOR BT RATING.

B. EXAMINATION OF AUTOCORRELATIONS AND PARTIAL AUTOCORRELATIONS

Autocorrelations describes the association between values of the same variable but at different time periods. Autocorrelation coefficients provide important information about the structure of a time series. These coefficients can be used to identify trends and possible seasonality within the data. [Ref. 4]

Partial autocorrelation is used to identify the extent of the relationship between current values of a variable with earlier values of that same variable, while holding the effects of all other time lags constant. [Ref. 4]

1. Yeoman

Examination of the autocorrelations for the Yeoman rating, Fig. 2.6, suggests that the data is stationary and should not require any transformation prior to model analysis. The residuals are within two standard errors of the mean zero and appear to be randomly distributed. The partial autocorrelations, Fig. 2.7, decay to zero rapidly and appear random after this point suggesting that an autoregressive model may be appropriate.

2. Storekeeper

Examination of the autocorrelations for this data set indicated the residuals met the randomness criteria of two standard errors. The shape of the residuals appeared to be a decaying sine curve Fig 2.8 The partial autocorrelations showed no shape but dropped to zero gradually and were randomly distributed, Fig. 2.9, not suggesting any clear cut model.

```
          -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
            '----'----'----'----'----'----'----'----'
 1   0.367                      XXXXXXXXXX
 2   0.171                      XXXXX
 3   0.144                      XXXXX
 4   0.100                      XXXX
 5   0.256                      XXXXXXX
 6   0.274                      XXXXXXXX
 7   0.014                      X
 8  -0.170                  XXXXX
 9  -0.092                    XXX
10  -0.004                      X
11   0.181                      XXXXXX
12   0.026                      XX
13  -0.050                     XX
14  -0.055                     XX
15   0.148                      XXXXX
```

Figure 2.6 AUTOCORRELATIONS FOR YN RATING.

```
          -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
            '----'----'----'----'----'----'----'----'
 1   0.367                      XXXXXXXXXX
 2   0.042                      XX
 3   0.080                      XXX
 4   0.022                      XX
 5   0.232                      XXXXXXX
 6   0.125                      XXXX
 7  -0.186                 XXXXXX
 8  -0.239                XXXXXXX
 9   0.004                      X
10   0.049                      XX
11   0.187                      XXXXXX
12  -0.110                   XXXX
13   0.050                      XX
14   0.042                      XX
15   0.243                      XXXXXXX
```

Figure 2.7 PARTIAL AUTOCORRELATIONS FOR YN RATING.

3. Operations Specialist

Analysis of the autocorrelations, Fig. 2.10, and partial autocorrelations, Fig. 2.11, for the Operations

```
              -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
                 I----I----I----I----I----I----I----I----I
     1    0.271                      XXXXXXXX
     2    0.160                      XXXXX
     3    0.192                      XXXXXX
     4   -0.024                     XX
     5   -0.025                     XX
     6    0.163                      XXXXX
     7    0.047                      XX
     8   -0.024                     XX
     9    0.054                      XX
    10    0.004                      X
    11    0.145                      XXXXX
    12   -0.000                      X
    13    0.002                      X
    14   -0.073                    XXX
    15   -0.026                     XX
```

Figure 2.8 AUTOCORRELATIONS FOR SK RATING.

```
              -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
                 I----I----I----I----I----I----I----I----I
     1    0.271                      XXXXXXXX
     2    0.093                      XXX
     3    0.139                      XXXX
     4   -0.128                   XXXX
     5   -0.027                     XX
     6    0.186                      XXXXXX
     7   -0.002                      X
     8   -0.077                    XXX
     9    0.010                      X
    10    0.023                      XX
    11    0.202                      XXXXXX
    12   -0.155                  XXXXX
    13   -0.025                     XX
    14   -0.100                    XXX
    15    0.092                      XXX
```

Figure 2.9 PARTIAL AUTOCORRELATIONS FOR SK RATING.

Specialist data set again indicated that the residuals were randomly distributed and had characteristics which strongly suggested the use of some type of autoregressive operation in model selection.

```
            -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
              '----'----'----'----'----'----'----'----'
 1   0.432                      XXXXXXXXXXXX
 2   0.157                      XXXXX
 3   0.047                      XX
 4  -0.088                    XXX
 5  -0.123                   XXXX
 6  -0.180                  XXXXX
 7  -0.257                XXXXXXX
 8  -0.262               XXXXXXXX
 9  -0.169                  XXXXX
10   0.038                      XX
11   0.285                      XXXXXXXX
12   0.220                      XXXXXXX
13   0.034                      XX
14  -0.032                     XX
15  -0.038                     XX
```

Figure 2.10 AUTOCORRELATIONS FOR OS RATING.

```
            -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
              '----'----'----'----'----'----'----'----'
 1   0.432                      XXXXXXXXXXXX
 2  -0.037                     XX
 3  -0.009                      X
 4  -0.123                   XXXX
 5  -0.045                     XX
 6  -0.119                   XXXX
 7  -0.159                  XXXXX
 8  -0.120                   XXXX
 9  -0.018                      X
10   0.139                      XXXX
11   0.257                      XXXXXXX
12  -0.040                     XX
13  -0.188                 XXXXXX
14  -0.112                   XXXX
15  -0.020                     XX
```

Figure 2.11 PARTIAL AUTOCORRELATIONS FOR OS RATING.

4. <u>Electronics Technicians</u>

The data set for the Electronics Technicians exhibited a strong trend when the residuals were evaluated, in

addition there was some non-stationarity suggested, which could require differencing[2] to remove. Fig. 2.12 illustrates the shape of the ACF function. Evaluation of the resultant autocorrelation showed that this process may not be necessary as the first lag exceeds -0.5 in magnitude which is a classic indication of an overdifferenced data set.

The shape of the autocorrelation and the partial autocorrelation, Fig. 2.13, suggest that an autoregressive or possibly a mixed model may be appropriate for evaluation.

```
               -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
                 I----I----I----I----I----I----I----I----I
  1    0.494                         XXXXXXXXXXXXX
  2    0.500                         XXXXXXXXXXXXX X
  3    0.429                         XXXXXXXXXXXX
  4    0.293                         XXXXXXXX
  5    0.336                         XXXXXXXXX
  6    0.134                         XXXX
  7    0.039                         XX
  8   -0.005                         X
  9   -0.030                        XX
 10   -0.140                     XXXXX
 11   -0.142                     XXXXX
 12   -0.140                     XXXXX
 13   -0.098                       XXX
 14   -0.126                      XXXX
 15    0.010                         X
```

Figure 2.12 AUTOCORRELATIONS FOR ET RATING.

---

[2] The method of differencing converts non-stationary time series into a stationary one. It consists of subtracting successive values from one another and using there difference as a new time series [Ref. 4].

```
             -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
               '----'----'----'----'----'----'----'----'
      1   0.494                    XXXXXXXXXXXXX
      2   0.339                    XXXXXXXXX
      3   0.147                    XXXXX
      4  -0.067                  XXX
      5   0.106                    XXXX
      6  -0.172                XXXXX
      7  -0.183               XXXXXX
      8  -0.068                  XXX
      9   0.062                    XXX
     10  -0.143                XXXXX
     11  -0.008                    X
     12   0.077                    XXX
     13   0.107                    XXXX
     14  -0.087                  XXX
     15   0.226                    XXXXXXX
```

Figure 2.13 PARTIAL AUTOCORRELATIONS FOR ET RATING.

5. Boiler Technicians

The Boiler technician rating data set exhibited no strong trend in the autocorrelations, Fig. 2.14. The shape of the autocorrelation and partial autocorrelation, Fig. 2.15, also strongly suggested that an autoregressive type of model should be considered for evaluation. This was indicated by the decaying sine wave pattern in the ACF and the abrupt cutoff of the value of the PACF.

C. MODEL DEVELOPMENT

In developing the models, the autocorrelations and partial autocorrelations were evaluated against representative Box-Jenkins models of the autoregressive (AR) and moving average (MA) type. A best fit model was then selected for evaluation utilizing the Minitab General Purpose Statistical Computing System. The results of these models were then evaluated to ensure the residuals were random and less than two standard errors of the mean zero. If the model

```
              -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
                +----+----+----+----+----+----+----+----+
  1   0.450                         XXXXXXXXXXXX
  2   0.208                         XXXXXX
  3  -0.028                        XX
  4  -0.238                   XXXXXXX
  5  -0.311                 XXXXXXXXX
  6  -0.394               XXXXXXXXXXX
  7  -0.360                XXXXXXXXXX
  8  -0.194                    XXXXXX
  9   0.101                         XXXX
 10   0.122                         XXXX
 11   0.180                         XXXXXX
 12   0.088                         XXX
 13  -0.009                         X
 14  -0.080                       XXX
 15  -0.117                      XXXX
```

Figure 2.14 AUTOCORRELATIONS FOR BT RATING.

```
              -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
                +----+----+----+----+----+----+----+----+
  1   0.450                         XXXXXXXXXXXX
  2   0.007                         X
  3  -0.155                     XXXXX
  4  -0.221                   XXXXXXX
  5  -0.133                      XXXX
  6  -0.221                   XXXXXXX
  7  -0.162                     XXXXX
  8  -0.028                        XX
  9   0.184                         XXXXXX
 10  -0.128                      XXXX
 11  -0.049                        XX
 12  -0.164                     XXXXX
 13  -0.140                      XXXX
 14  -0.144                     XXXXX
```

Figure 2.15 PARTIAL AUTOCORRELATIONS FOR BT RATING.

was satisfactory at this point, the sum of squared errors (SSE) was evaluated to determine if it was the lowest possible reduction in the SSE. In the event that more than one model passed these tests the t-ratios were then

evaluated. Again, if all these tests were insignificantly different among a family of models the principle of parsimony was used to select the "best" model. In every model developed during this phase of analysis, all three of the criteria were met by only the model selected which made the choice of the correct prediction model relatively easy to select.

Once these models were selected, forecasts were generated for the period October 1983 to March 1984 and compared to the actual reenlistment totals for the period. If a model was evaluated as totally inappropriate at this point, then further investigation and modeling was pursued to attempt resolution of the problem.

Table II presents a summary of the model forecasts for the selected ratings. Model summaries and statistics are presented in appendix C.

TABLE II

RESULTS OF UNIVARIATE MODEL FORECASTS

PERIOD	FORECAST	95% FORECAST LIMITS LOWER	UPPER	ACTUAL	ERROR
YN RATING					
OCT 83	67.9	46.3	89.5	60.7	11.8
NOV 83	61.3	37.6	84.9	55.6	10.1
DEC 83	58.3	34.2	82.4	66.0	11.6
JAN 84	57.0	32.8	81.1	60.8	17.2
FEB 84	56.4	32.2	80.5	59.1	04.6
MAR 84	56.1	31.9	80.2	51.8	08.2
SK RATING					
OCT 83	60.4	35.6	85.2	64.6	06.4
NOV 83	60.4	35.1	85.6	69.8	13.4
DEC 83	60.4	34.7	86.1	70.8	14.6
JAN 84	60.4	34.3	86.5	72.0	16.0
FEB 84	60.4	33.9	86.9	66.7	09.4
MAR 84	60.4	33.5	87.3	75.0	19.4
OS RATING					
OCT 83	44.6	13.4	75.8	41.5	07.5
NOV 83	43.7	08.8	78.5	41.2	06.0
DEC 83	43.2	07.4	78.9	52.6	17.8
JAN 84	43.0	07.0	78.9	52.7	18.4
FEB 84	42.9	06.9	78.9	48.9	12.3
MAR 84	42.8	06.8	78.8	62.5	31.5
ET RATING					
OCT 83	95.3	84.7	100	95.9	00.6
NOV 83	93.6	81.3	100	95.6	02.0
DEC 83	92.6	79.8	100	95.5	00.3
JAN 84	92.1	79.0	100	96.2	04.3
FEB 84	91.7	78.6	100	94.2	02.6
MAR 84	91.5	78.4	100	95.9	04.5
BT RATING					
OCT 83	52.5	27.4	77.6	42.6	23.2
NOV 83	47.2	18.7	75.8	56.5	16.4
DEC 83	44.4	14.9	73.9	48.0	07.5
JAN 84	42.8	13.0	72.6	52.0	17.6
FEB 84	42.0	12.1	71.8	55.6	24.5
MAR 84	41.5	11.7	71.4	61.8	32.8

1. Yeoman Model

The model selected for the Yeoman rating was of the ARIMA (1, 0, 0) type, a comparison of the forecasted re-enlistment percentages with the actual totals showed an average error of .106 with a range from a low of -.172 to a high of .118. all of the observations were within the 95% confidence limits of the model and were also captured in shape by the model. Fig. 2.16 graphically illustrates the forecasts and observations.

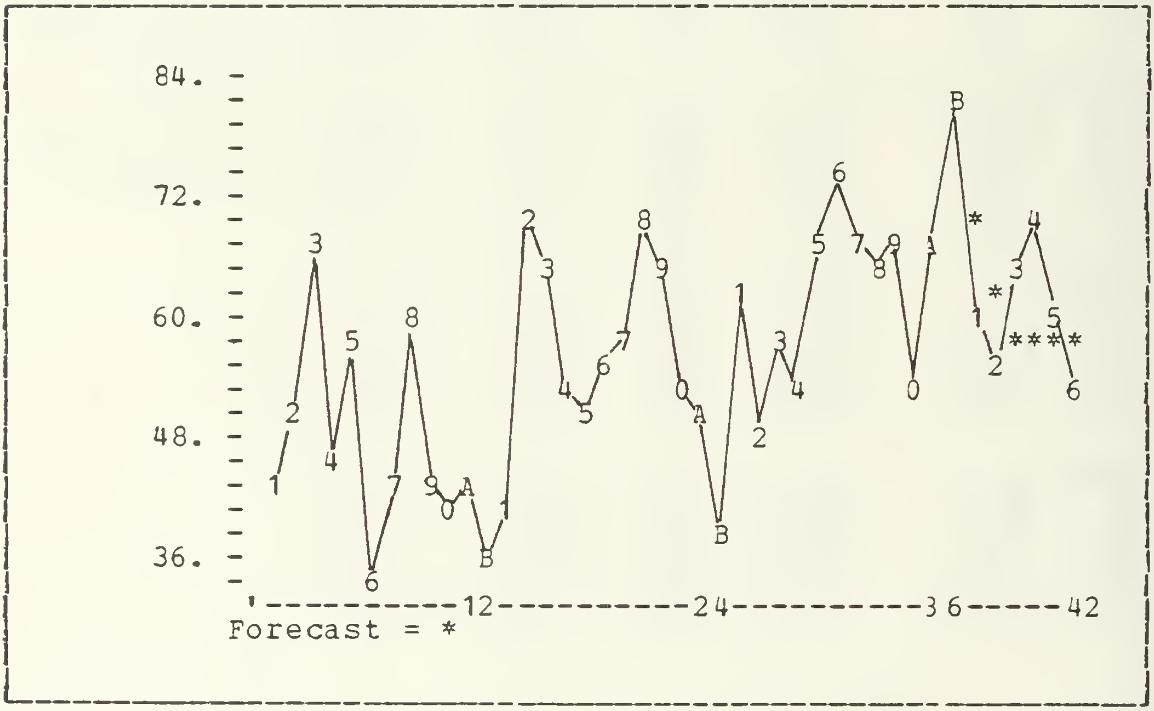

Figure 2.16 PLOT OF UNIVARIATE MODEL FOR YN RATING.

2. Storekeeper Model

Since the autocorrelations and partial autocorrelations did not clearly suggest the selection of one model type as being superior to another, an iterative method was used for selection. After discarding several first order AR, MA and ARMA models it was decided to try differencing even though the data did not strongly suggest that this was necessary. The resultant ACF and PACF satisfied the randomness criteria without indicating overdifferencing. A model of the ARIMA (0,1,1) order was then found to meet all the necesarry stringency requirements for model selection. The model generated forecasts with an average error of .132 and a range from -.064 to -.194. It is felt that iterative modeling with the actual data for the forecast period would reduce this error even further. Fig. 2.17 illustrates the data fit with the model.

3. Operations Specialist Model

Modeling of the Operations Specialist rating resulted in the selection of an ARIMA (1,0,0) as the most appropriate model. The data for the rating presented the largest fluctuation in range over the entire data set. These fluctuations have no doubt influenced the ultimate predictive power of the selected model. As a result, the model yielded acceptable predictions varying from the actual observations by an average of .155 and a range from .06 to -.315. As with the previous model, successive iterations with the new observations should improve the predictive power of the model. Fig. 2.18 illustrates the models fit with the actual observations.

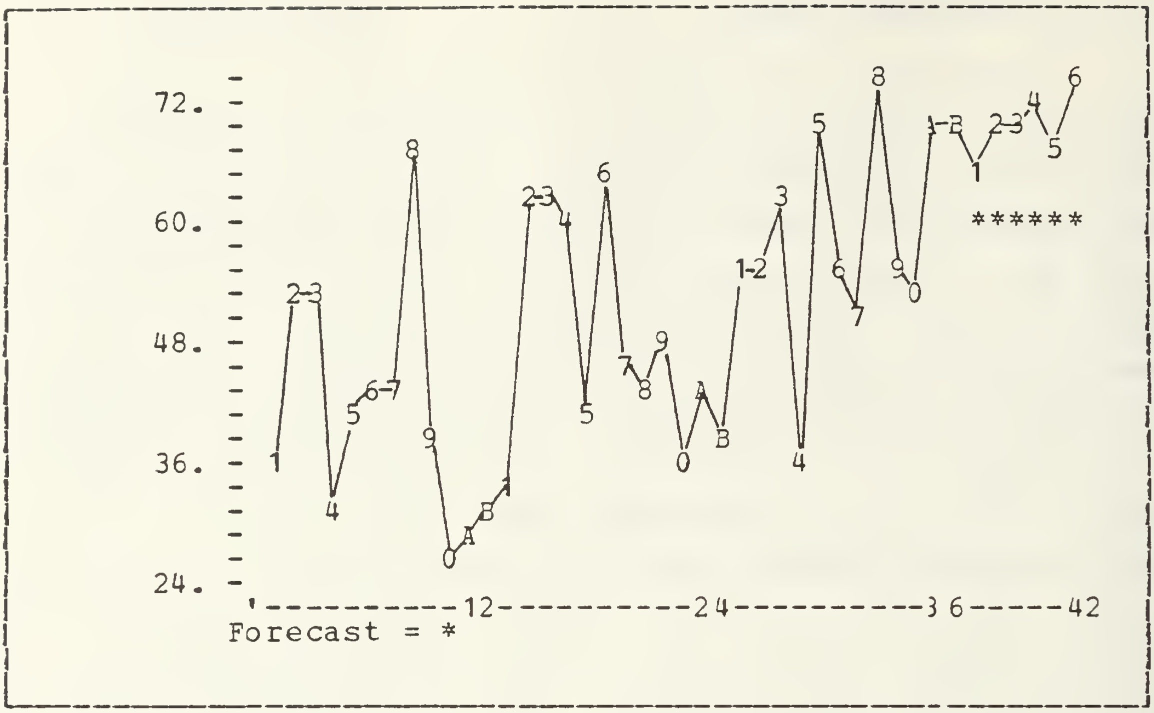

Figure 2.17 PLOT OF UNIVARIATE MODEL FOR SK RATING.

4. Electronics Technician Model

A first order autoregressive model of the ARIMA (1,0,0) type was also selected for the electronics technician rating. The model produced spectacular results with an average error of -.025 and a range of errors from -.003 to -.045. It should be noted that this data set is also the most stable over time with an average of more than 91% first term reenlistments. While this makes the model's job somewhat easier, it still remains as a powerful model for univariate predictions. Fig. 2.19 illustrates the fitting of the model.

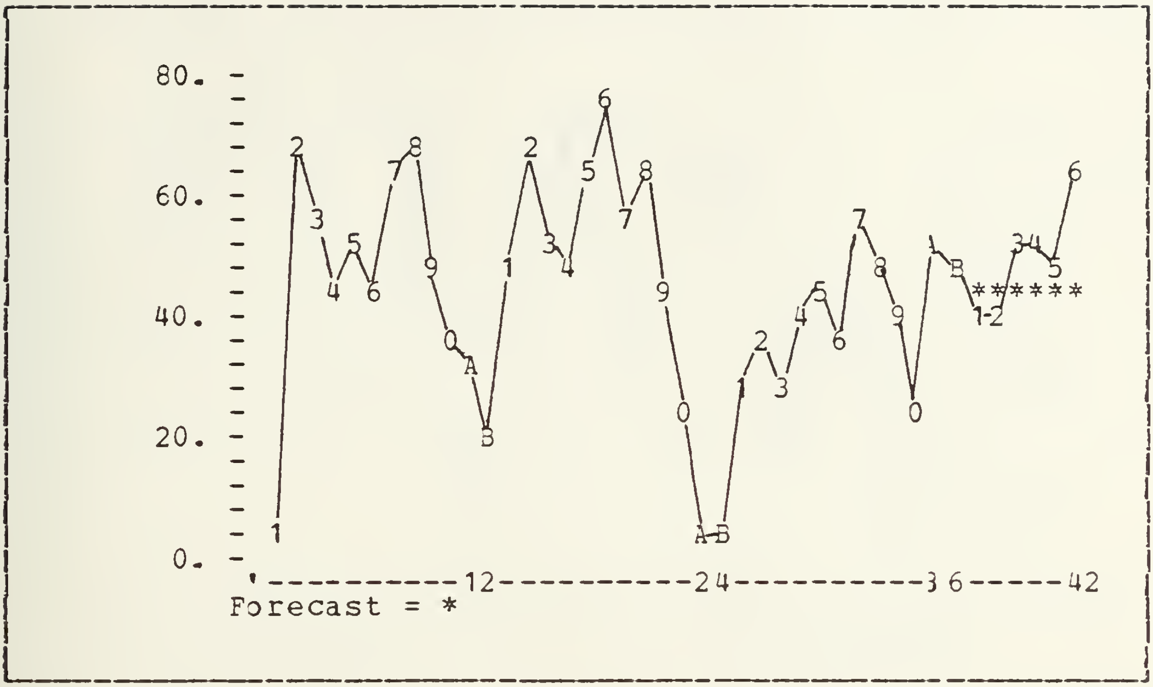

Figure 2.18 PLOT OF UNIVARIATE MODEL FOR OS RATING.

5. Boiler Technician Model

In the case of the boiler technicians, an ARIMA (1,0,0) was again evaluated as the most appropriate, however; this model was the poorest predictor of any selected for evaluation. The average error was .203 with a range from -.328 to .232. The model also failed to capture the shape of the actual data which presented an upward trend while the model indicated a downward turn. The model is however acceptable for further analysis and refinement in the transfer function model. Fig. 2.20 illustrates the fitting of the data set predictions.

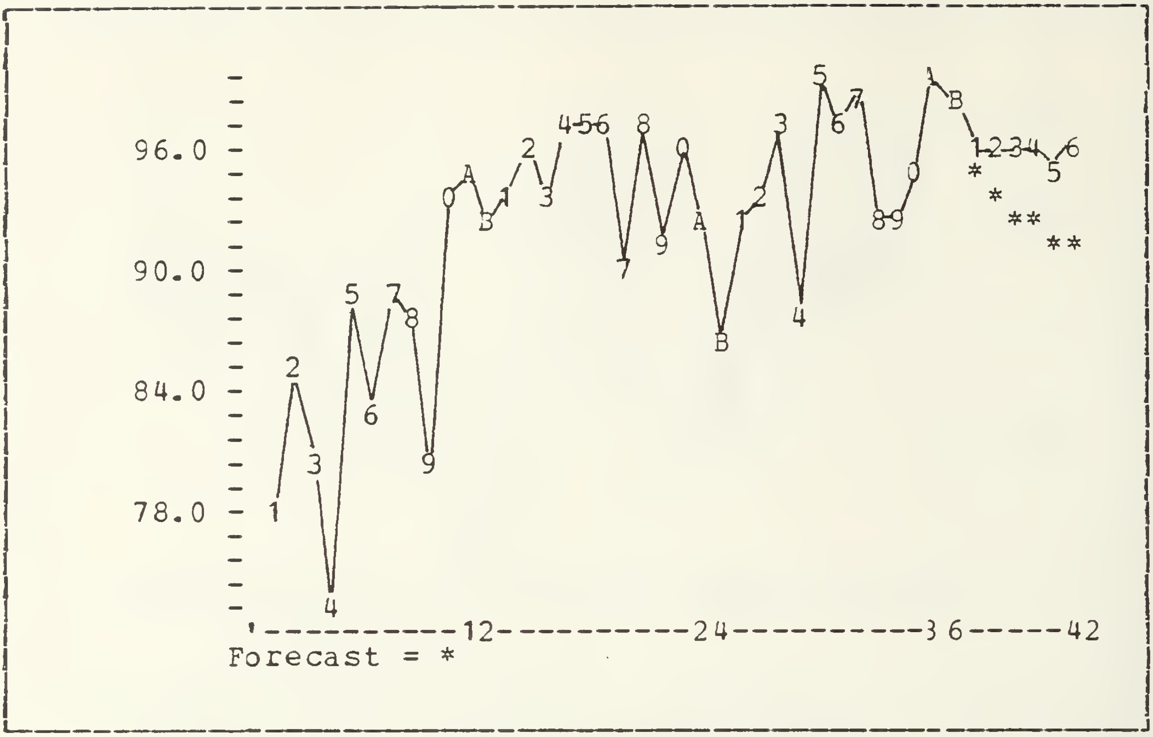

Figure 2.19 PLOT OF UNIVARIATE MODEL FOR ET RATING.

D. SUMMARY

In this chapter, five acceptable univariate models have been developed for the respective data sets. These models were in most cases fairly obvious from the analysis of the autocorrelations and partial autocorrelations, however the process can be fairly time consuming and result in the pursuit of several "blind alleys" on the way to a workable model. In succeeding chapters, a transfer function model utilizing unemployment statistics for the 20-24 age group will be developed.

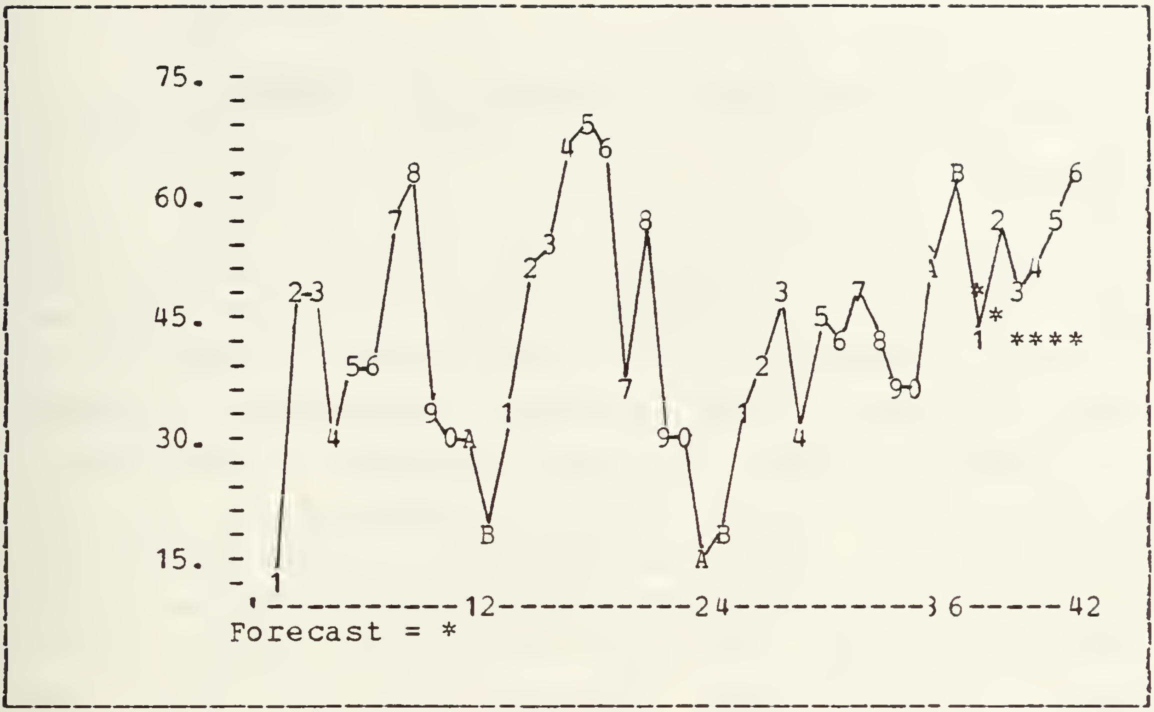

Figure 2.20 PLOT OF UNIVARIATE MODEL FOR BT RATING.

III. THE LEADING INDICATOR, REGRESSION AND COMBINED MODELS

A. OVERVIEW

A leading indicator model for reenlistment in the selected ratings was constructed by first developing a univariate time series model for unemployment in the 20-24 year old age group and then applying this model to the data for the selected ratings. The resultant model residuals were then crosscorrelated to establish the location of any time leads or lags that affect reenlistments. These indicators could provide an early warning system of shifts in the level of reenlistment and/or the direction of the trend in reenlistment. Once developed, the adequacy of the model was tested by using the coefficient of determination (R-squared) from the indicated lag/lead. Forecasts for the Oct 83-Mar 84 time period were also generated in this process and the results compared to the univariate models forecasts. A combined model using time series analysis and regression were also formulated.

B. THE UNEMPLOYMENT MODEL

The data for unemployment in the 20-24 age bracket was collected from monthly publications by the Bureau of Labor Statistics. These figures cover the period from October 1980 through September 1983 and are presented in Appendix A. This particular age grouping was selected as being the most appropriate for personnel completing their first enlistment and facing the reenlistment decision.

1. Methodology

The unemployment time series model was constructed utilizing the same methodology as applied in the preceding chapter to reenlistment time series for the selected rating models. The unemployment data was initially transformed by computing the relative percentage change from one period to the next. This was done by subtracting the rate in the current period from the rate in the preceding period and dividing the remainder by the rate of the preceding period. In doing this, it was felt that the resulting model would better capture responses to changes in the unemployment rate rather than responses to the overall level. It was further hypothesized that this would capture any perceptions by the service member that the job market was improving or worsening in relation to the demand for a particular skill [Ref. 5].

The computed change in unemployment time series was then evaluated for a potential model by screening the autocorrelation and partial autocorrelation functions. The data appeared stationary but did not suggest any obvious model for selection. As a result a trial and error method was used for model selection. The model iterations were evaluated for suitability utilizing the same criteria described in the previous chapter, that is evaluation of residuals for randomness, smallest sum of squared errors and a *t*-ratio in excess of 2.0. Several trials of autoregressive models, moving average models and mixed autoregressive moving average models did not yield any positive results. Therefore, it was decided to difference the data one time and try the iterative model building process again.

The resulting differenced data set also met the stationarity criteria and did not exhibit any characteristics of an overdifferenced data set. When evaluated at this

point, the autocorrelation and partial autocorrelation suggested that a moving average model was appropriate. This model of the ARIMA (0,1,1) type met all of the selection criteria necessarry and was therefore adopted for use. Model results and specifications are presented in Fig. 3.1 through Fig. 3.4.

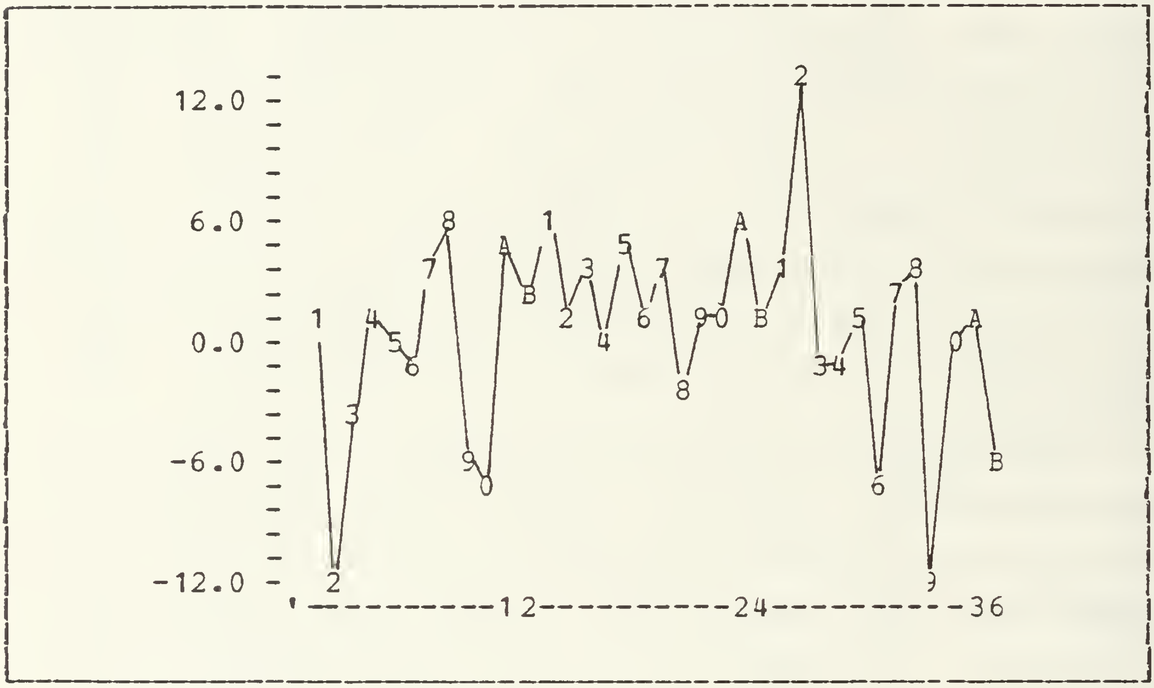

Figure 3.1 TIME SERIES PLOT FOR UNEMPLOYMENT DATA.

```
                  -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
                    |----|----|----|----|----|----|----|----|
  1   0.025                             XX
  2  -0.115                          XXXX
  3   0.228                             XXXXXXX
  4  -0.069                           XXX
  5  -0.083                           XXX
  6   0.090                             XXX
  7  -0.045                            XX
  8   0.063                             XXX
  9   0.002                             X
 10  -0.124                          XXXX
 11  -0.038                            XX
 12   0.011                             X
 13  -0.032                            XX
 14  -0.034                            XX
 15  -0.029                            XX
```

Figure 3.2 AUTOCORRELATION FUNCTION FOR UNEMPLOYMENT DATA.

```
                  -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
                    |----|----|----|----|----|----|----|----|
  1   0.025                             XX
  2  -0.115                          XXXX
  3   0.238                             XXXXXXX
  4  -0.109                          XXXX
  5  -0.017                             X
  6   0.022                             XX
  7  -0.028                            XX
  8   0.109                             XXXX
  9  -0.063                           XXX
 10  -0.080                           XXX
 11  -0.074                           XXX
 12   0.012                             X
 13   0.015                             X
 14  -0.045                            XX
 15  -0.043                            XX
```

Figure 3.3 PARTIAL AUTOCORRELATIONS FOR UNEMPLOYMENT DATA.

TABLE III

SUMMARY OF UNEMPLOYMENT MA 1 MODEL

NUMBER	TYPE	ESTIMATE	ST. DEV.	T-RATIO
1	MA 1	0.9850	0.0689	14.29

```
 DIFFERENCING. 1 REGULAR
 RESIDUALS.   SS = 896.9  (BACKFORECASTS EXCLUDED)
              DF = 34  MS = 26.4
 NO. OF OBS. ORIGINAL SERIES 36 AFTER DIFFERENCING 35
```

```
              -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
                 .    .    .    .    .    .    .    .    .
  1   0.041                          XX
  2  -0.111                       XXXX
  3   0.227                          XXXXXXX
  4  -0.068                        XXX
  5  -0.081                        XXX
  6   0.087                          XXX
  7  -0.051                         XX
  8   0.074                          XXX
  9   0.012                          X
 10  -0.127                       XXXX
 11  -0.040                         XX
 12   0.004                          X
 13  -0.033                         XX
 14  -0.038                         XX
 15  -0.029                         XX
```

Figure 3.4 AUTOCORRELATION OF RESIDUALS FOR UNEMPLOYMENT MODEL.

C. APPLICATION TO THE SELECTED RATINGS DATA SETS

1. Yeoman

When the ARIMA (0,1,1) model was applied to both the reenlistment data for yeoman and the differenced unemployment data set and the cross-correlation function evaluated, there appeared to be a relationship at the twelve month lead point. The value at this point was significant when compared to the other points however, it was not significant in absolute terms since it was not in excess of two standard errors of the mean zero. This observation of magnitude holds true in all of the models. Fig. 3.5 shows the cross-correlation function for this data set for all lead months only.

```
              -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
                '----'----'----'----'----'----'----'----'
-15  -0.110                      XXXX
-14  -0.044                        XX
-13  -0.204                    XXXXXX
-12   0.191                         XXXXXX
-11   0.115                         XXXX
-10   0.073                         XXX
 -9   0.072                         XXX
 -8  -0.023                        XX
 -7  -0.106                      XXXX
 -6  -0.038                        XX
 -5   0.053                         XX
 -4  -0.041                        XX
 -3  -0.030                        XX
 -2  -0.200                    XXXXXX
 -1  -0.155                     XXXXX
  0  -0.135                      XXXX
```

Figure 3.5 CROSS CORRELATION FOR YN RATING.

2. Storekeepers

The cross-correlation function for the storekeeper model indicated a possible lead indicator relationship at the five month and eleven month points. Both of these points were significant in their relationship to the other values but again were below the accepted level for determining significance. It was decided to investigate the significance of these points in the regression procedure. Fig. 3.6 shows the cross-correlation function for positive leads of this model.

```
           -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
             '----'----'----'----'----'----'----'----'
-15  -0.211                 XXXXXX
-14  -0.069                    XXX
-13  -0.063                    XXX
-12  -0.006                      X
-11   0.213                      XXXXXX
-10   0.120                      XXXX
 -9   0.016                      X
 -8   0.060                      XXX
 -7  -0.162                  XXXXX
 -6  -0.102                   XXXX
 -5   0.218                      XXXXXX
 -4  -0.374             XXXXXXXXXX
 -3  -0.107                   XXXX
 -2  -0.154                  XXXXX
 -1  -0.290               XXXXXXXX
  0   0.140                      XXXXX
```

Figure 3.6 CROSS CORRELATION FOR SK RATING.

3. Operations Specialists

The cross-correlation function again indicated more than one point for possible investigation as being relatively significant. These points occurred at the six, nine and twelve month points. Fig. 3.7 again shows the lead values for this model.

```
              -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
                 +----+----+----+----+----+----+----+----+
 -15    0.006                        X
 -14    0.024                        XX
 -13    0.071                        XXX
 -12    0.200                        XXXXXX
 -11    0.170                        XXXXX
 -10    0.102                        XXXX
  -9    0.192                        XXXXXX
  -8   -0.140                    XXXXX
  -7   -0.240                  XXXXXXX
  -6    0.245                        XXXXXXX
  -5   -0.034                       XX
  -4   -0.208                   XXXXXX
  -3   -0.165                    XXXXX
  -2   -0.278                 XXXXXXXX
  -1   -0.038                       XX
   0   -0.019                        X
```

Figure 3.7 CROSS CORRELATION FOR OS RATING.

4. Electronics Technicians

The cross-correlation function for this model also suggested more than one lead point for investigation, these occurred at the six and twelve month points but again were significant only in relative terms and not in absolute magnitude. Fig. 3.8 illustrates the lead relationship in the cross-correlation function.

```
             -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
               +----+----+----+----+----+----+----+----+
  -15   0.138                      XXXX
  -14   0.098                      XXX
  -13  -0.003                      X
  -12   0.275                      XXXXXXXX
  -11   0.031                      XX
  -10   0.044                      XX
   -9   0.179                      XXXXX
   -8   0.014                      X
   -7  -0.127                   XXXX
   -6   0.248                      XXXXXXX
   -5   0.091                      XXX
   -4  -0.223                XXXXXXX
   -3   0.038                      XX
   -2  -0.114                   XXXX
   -1   0.122                      XXXX
    0  -0.017                      X
```

Figure 3.8 CROSS CORRELATION FOR ET RATING.

5. Boiler Technicians

Consistent with the previous models, tie model for Boiler Technicians also presented more than one prominent point for evaluation. These points occurred at the nine and twelve month points. Fig. 3.9 is the cross-correlation function for the lead values of the model.

```
                  -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
                    I----I----I----I----I----I----I----I----I
 -15   0.084                            XXX
 -14  -0.011                            X
 -13  -0.072                          XXX
 -12   0.262                            XXXXXXXX
 -11   0.190                            XXXXXX
 -10   0.133                            XXXX
  -9   0.160                            XXXXX
  -8  -0.096                          XXX
  -7  -0.294                     XXXXXXXX
  -6   0.141                            XXXXX
  -5   0.034                            XX
  -4  -0.286                     XXXXXXXX
  -3  -0.154                        XXXXX
  -2  -0.159                        XXXXX
  -1   0.088                            XXX
   0  -0.007                            X
```

Figure 3.9 CROSS CORRELATION FOR BT RATING.

D. LINEAR REGRESSION

In order to verify the leading indicator models, a linear regression will be constructed using 20-24 year old unemployment as the independent variable and reenlistment rates for the selected ratings as the dependent variable. For this process to be valid, certain assumptions are required prior to application of the process. The first assumption is that there is a linear relationship between the variables as described by eqn. 3.1

$$Y = a + BX_i + e \qquad \text{(eqn 3.1)}$$

where for each observation, Y is a random variable. The second assumption is that X is fixed in value and the final assumption is that e, the error term, has an expected value of zero with constant variance for all observations. It is further assumed that the e's are normally distributed and uncorrelated. [Ref. 6]

E. MODEL VERIFICATION

Linear regression models were run for all of the data sets using the lagged value of the unemployment data as the independent variable and the reenlistment data as the dependent variable. The regressions were evaluated using the R-squared value, Durbin-Watson statistic and <u>t</u>-ratio A summary of the regression models R-squared and Durbin Watson statistic is presented in Table IV.

The regression models for all of the ratings were less than robust in their ability to verify the significance of the leading indicator models. R-squared values ranged from a high of .34 for the Yeoman model to a dismal -.02 for the

Boiler Technician model. These values represent the best R-squared that were obtained at all lagged values of the independent variable and not just the ones that were indicated as being significant in the leading indicator model. _t_-ratios were somewhat stronger ranging from .1 for Storekeepers to 11.3 for Electronics Technicians with three of the five ratings having values above the minimum acceptance level of 2.0. The Durbin-Watson statistic was weak for the models as well with only the Storekeeper model clearly within the acceptable range of 1.5 to 2.5 for the data. This indicates that the regression failed to remove all of the serial correlation present in the data sets and the residuals are positively correlated to one another.

While these models are disappointing, they are not discouraging. They seem to indicate that, when taken alone, unemployment does not possess the strong predictive ability it seems to have in other econometric models. Bepko [Ref. 1], and Darling [Ref. 2], in their regression models attribute nearly 50 percent of the explained sum of squared error to unemployment, this may indicate that this relationship may not hold in modeling the behavior of first term reenlistments for military personnel. It should be noted that Bepko [Ref. 1], constructed an aggregated model of careerists using the 25-39 age group for unemployment and Darling [Ref. 2], utilized national teenage unemployment in modeling Marine Corp enlistments.

The predictions for the regression model are compared to the observed levels of reenlistment for the period October 1983 to March 1984, these forecasted levels are shown in table VI In addition, the regression forecast for the entire data set from October 1980 to September 1983 will be generated and the residuals for that data set will be independently modeled as a new time series with a forecast of residuals to be applied to the forecasts from the regression

TABLE IV

SUMMARY OF R-SQUARED AND DURBIN-WATSON STATISTICS

X=UNEMPLOYMENT FOR 20-24 YEAR OLD AGE GROUP
SELECTED RATINGS R-SQUARED VALUES

LAGS (X)	(DURBIN-WATSON STATISTIC IN PARENS)				
	YN	SK	OS	ET	BT
1	.265 (1.54)	.22 (1.77)	.027 (.82)	.181 (1.22)	-.03 (.87)
2	.272 (1.56)	.154 (1.81)	.078 (.83)	.141 (1.2)	-.03 (.87)
3	.30 (1.55)	.199 (1.56)	.025 (.84)	.133 (1.22)	-.031 (.84)
4	.349 (1.43)	.163 (1.72)	.037 (.81)	.115 (1.51)	-.027 (.83)
5	.213 (1.3)	.134 (1.76)	.031 (.81)	.041 (1.39)	-.023 (.84)
6	.259 (1.31)	.229 (1.76)	.006 (.79)	.022 (1.55)	-.032 (.79)
7	.272 (1.25)	.253 (1.75)	.002 (.8)	.003 (1.6)	-.035 (.81)
8	.21 (1.16)	.224 (1.6)	.004 (.79)	-.009 (1.65)	-.034 (.74)
9	.148 (1.16)	.183 (1.58)	-.023 (.81)	.024 (2.08)	-.027 (.73)
10	.171 (1.14)	.193 (1.69)	-.023 (.79)	.024 (2.08)	-.027 (.75)
11	.098 (1.3)	.115 (1.97)	-.013 (.74)	-.011 (2.08)	-.043 (.75)
12	.203 (1.66)	.130 (2.24)	-.045 (.71)	.051 (2.09)	-.042 (.8)

models. This procedure should yield forecasts with a better fit than was possible with only the univariate or regression models. This procedure closely parallels the work of Darling [Ref. 2], in modeling the supply of recruits for the Marine Corp. The combined model will be developed in the next section.

F. THE COMBINED MODEL

The regression models in the previous section yielded adequate forecasts of the reenlistment rate, they were not, however, significantly better than the forecasts for the univariate models in chapter three. Additionally, the residuals of the regression model exhibited strong positive serial correlation as indicated by the low values of the Durbin-Watson statistic for each model, Table IV summarizes the data for the regression models. As shown by Darling [Ref. 2], this enables the residuals to be constructed into an independent time series. Through the application of the Box-Jenkins method a forecast of the residuals or error terms can be generated and applied to the regression models forecasts. This procedure should yield a forecast with a better fit to the actual data.

TABLE V

REGRESSION MODELS FOR SELECTED RATINGS

RATE	LAG	CONSTANT	BETA	ST.DEV.	R-SQRD	DURBIN-WATSON
YN	4	9.05	3.32	9.97	.327	1.43
SK	7	1.59	3.57	11.88	.253	1.75
OS	4	73.2	-2.08	16.82	.037	.81
ET	1	73.2	1.31	5.53	.181	1.22
BT	5	52.6	-.743	14.93	-.023	.84

The Box-Jenkins methodology as described in Chapter three was again applied to the sets of regression model residuals and appropriate models were selected. Table VI is

a summary of the forecasts for all three methods and percentage of error for each forecast.

The combined model resulted in improved forecasts in three of the five ratings with the other two showing either minor improvement or a slight decline (BT, OS). It should be noted at this point that these two data sets were the most volatile in terms of range of observations. The forecasts for these ratings could be improved by eliminating significant outliers from the data sets and recomputing all of the models. Due to the already small size of the data sets, this was not considered. This problem should be corrected in future works in this area as more data points become available for analysis.

TABLE VI

FORECAST COMPARISONS OF ALL THREE MODELS

DATE	ACTUAL RATE OF REENLISTMENTS	BOX-JENKINS FORECASTS	REGRESSION FORECASTS	COMBINED BOX-JENKINS REGRESSION FORECASTS
YEOMAN				
OCT 83	60.7	67.9 (11.8)	51.2 (.8)	66.3 (9.2)
NOV 83	55.6	61.3 (10.1)	51.2 (10.1)	61.4 (10.4)
DEC 83	66.0	58.3 (-11.6)	51.8 (-6.4)	62.0 (-6.0)
JAN 84	68.8	57.0 (-17.2)	58.9 (-14.4)	59.1 (-14.1)
FEB 84	59.1	56.4 (-4.6)	58.2 (-.5)	58.4 (-1.2)
MAR 84	51.8	56.1 (8.2)	54.5 (5.2)	54.7 (5.6)
STOREKEEPER				
OCT 83	64.6	60.4 (-6.4)	50.9 (-5.7)	65.6 (1.5)
NOV 83	69.8	60.4 (-13.4)	52.3 (-10.7)	66.5 (-4.7)
DEC 83	70.8	60.4 (-14.6)	54.4 (-9.0)	68.9 (-2.7)
JAN 84	72.0	60.4 (-16.0)	57.6 (-20.0)	62.0 (-13.9)
FEB 84	66.7	60.4 (-9.4)	57.6 (-13.6)	62.0 (-7.0)
MAR 84	75.0	60.4 (-19.4)	58.4 (-22.1)	62.8 (-16.2)
OPERATION SPECIALIST				
OCT 83	41.5	44.6 (7.5)	40.5 (-2.4)	47.3 (14.0)
NOV 83	41.2	43.7 (6.0)	40.5 (-1.6)	45.5 (10.4)
DEC 83	52.6	43.2 (-17.8)	40.1 (-23.8)	44.0 (-16.4)
JAN 84	52.7	43.0 (-18.4)	42.0 (-20.3)	45.2 (-14.2)
FEB 84	48.9	42.9 (-12.3)	42.4 (-13.3)	45.2 (-7.6)
MAR 84	62.5	42.8 (-31.5)	44.7 (-28.5)	47.3 (-24.4)
ELECTRONICS TECHNICIANS				
OCT 83	95.9	95.3 (-.6)	92.9 (-3.1)	95.1 (-.8)
NOV 83	95.6	93.6 (-2.0)	92.6 (-3.1)	94.4 (-1.3)
DEC 83	95.5	92.6 (-3.0)	91.2 (-4.5)	92.7 (-2.9)
JAN 84	96.2	92.1 (-4.3)	92.5 (-3.8)	93.7 (-2.3)
FEB 84	94.2	91.7 (-2.6)	92.5 (-1.8)	93.5 (-.7)
MAR 84	95.9	91.5 (-4.5)	92.6 (-3.4)	93.4 (-2.6)
BOILER TECHNICIANS				
OCT 83	42.6	52.5 (23.2)	39.5 (-7.2)	53.0 (24.4)
NOV 83	56.5	47.2 (-16.4)	40.9 (-27.4)	49.2 (-13.3)
DEC 83	48.0	44.4 (-7.5)	40.9 (-14.8)	46.1 (-3.9)
JAN 84	52.0	42.8 (-17.6)	40.8 (-21.5)	44.2 (15.0)
FEB 84	55.6	42.0 (-24.5)	41.5 (-25.4)	43.9 (-21.0)
MAR 84	61.8	41.5 (-32.8)	41.6 (-32.7)	43.4 (-29.7)

IV. SUMMARY AND CONCLUSIONS

A. SUMMARY

Several forecasting techniques have been examined in this thesis in an attempt to predict the pattern of reenlistments in five specific ratings. Two distinct methods were used to build three models; a univariate Box-Jenkins model, a linear regression model and a combined regression and Box-Jenkins model.

The results of each varied in predictive ability, with the combined model being clearly superior to the other two, (as measured by percent error of the actual observations). but with the results by rating differed sharply within each model. For electronics technicians all three models were clearly adequate, this is not surprising since this rating had the smallest range and the least variance in reenlistments during the time period examined. The regression equation for ET's yielded a very low R-squared value of the model. Appropriate additional explanatory variables may be the level of reenlistment bonuses or the availability of advanced technical training. Boiler technicians and operations specialists showed the widest range in reenlistment percentages, and, as expected, their models exhibited the least accuracy. The regression equations for these ratings were counterintuitive in that they indicated higher reenlistment rates at successively lower levels of the independent variable, unemployment. This indicates that an additional independent variable may be required in the equation for these ratings. For BT's this may be a dummy variable accounting for the unpleasantness of the working

conditions or the level of their reenlistment bonuses. For the OS rating, it may also be the reenlistment bonus level or a factor accounting for the high amount of sea duty present in that rating when compared to others. The ratings of yeoman and storekeeper presented models that were marginal when using Box-Jenkins or regression separately but presented quite good predictions when utilizing the combined model.

A somewhat surprising result of the models forecasts was the accuracy of the regression model using 20-24 year old unemployment as the only independent variable. This is surprising in view of the low R-squared values of the models and the high degree of serial correllation remaining in the residuals as expressed by the Durbin-Watson statistic. This was actually the second most accurate prediction model outperforming the univariate Box-Jenkins model by a slight margin.

The Box-Jenkins models' performance was restricted by the size of the data set available. Technically, thirty or more observations in a data set are considered sufficient but 100 or more observations are considered desirable in order to utilize the full predictive power of the model. This larger number of observations is also considered desirable in terms of identifying the underlying trends and patterns which may not appear in a smaller set of data. In terms of forecasting reenlistments, it is not possible at this point to utilize any more data points than were available for this study since the monthly figures are aggregated by quarters after three years and only the quarterly data are retained.

B. CONCLUSIONS

A surprising finding, for all of the ratings modeled, is the continued rise in reenlistments in view of the ever improving economy during the period. This could possibly be explained, in a regression model, with the introduction of the civilian/military pay ratio for the period or the level of reenlistment bonuses for a rating. This would still, however, not account for the 95 percent reenlistment rate for electronics technicians who are generally regarded as having the most desireable and marketable skills in almost any employment market. Another explanatory term could be introduced for "taste" for military service much as the ACOL model uses. In light of the world situation and recent events in Lebanon, Granada and the Persian Gulf this could be a significant explanatory factor for the continued rise in reenlistments.

In terms of policy implications, the results for all of the models utilized indicate that high levels of first term retention are likely to continue in all of these ratings for the next six to twelve months. At this point, decisions will be required on how to deal with these increases in a service that is rapidly approaching authorized end strength. The longer range forecasts still seem to indicate that the currently favorable climate will eventually give way to an ever improving economy. Now would seem to be the most opportune time to take advantage of the situation by increasing the total number of personnel in the career force as a hedge against the future change in the demographics of the cohort eligible for military service. This will induce a short term increase in compensation costs by increasing the inventory of career petty officers. This will eventually be offset by the reduction in future training and recruiting costs that will result from this larger career force.

C. SUGGESTIONS FOR FUTURE RESEARCH

As previously stated, the full potential of the Box-Jenkins method has not been fully exploited because of restrictions in the amount of data available. Further, this restriction can only be corrected with the passage of time as more observations become available. The models presented in this paper were rating-specific, which may only have limited application. In a broad sense, however, research should continue along these lines with aggregate models of rating groups. As to how this aggregation should be performed, Thomas and Liao [Ref. 3], have suggested grouping ratings by observed reenlistment behavior in their model of second and subsequent term careerists. This grouping should be conducive to application of Box-Jenkins techniques which appears to be more effective when dealing with a data set of narrow range. Another possibility for grouping could follow the level of skill required as indicated by a rating being termed high-tech, medium tech or low tech [Ref. 1].

The combined regression, Box-Jenkins model presented here also deserves future consideration as it appears to be a viable "fine tuning" method for regression models and intuitively more appealing than introducing more and more variables into the analysis. Use of a combined modeling technique can only serve to strengthen the results of regression models that are currently very popular.

The Box-Jenkins method is not meant to be an all encompassing method for use in manpower modeling. It certainly should, however, be considered as a tool to be placed in the arsenal of the manpower planner for continued use and development. In view of the many commercial software packages available for this technique, implementation and application to manpower issues should most strongly be considered for use in Navy manpower planning.

APPENDIX A

SUMMARY OF DATA USED IN ANALYSIS

TABLE VII

DATA UTILIZED FOR ANALYSIS

Rating Date	YN	SK	OS	ET	BT	Unemployment
Oct 80	43.2	36.2	5.1	77.7	12.2	13.8
Nov 80	50.	53.4	68.1	85.1	46.9	12.1
Dec 80	66.1	53.6	55.	80.2	47.4	11.7
Jan 81	46.	30.3	43.	73.6	29.8	11.8
Feb 81	58.6	41.6	51.	88.6	39.9	11.8
Mar 81	34.	43.7	45.3	82.2	39.4	11.7
Apr 81	43.2	44.4	63.2	88.5	56.6	12.1
May 81	60.	67.2	68.1	87.1	62.8	12.9
Jun 81	44.1	37.3	49.1	80.8	33.7	12.1
Jul 81	40.7	26.9	37.3	93.7	29.8	11.3
Aug 81	42.5	28.7	30.6	94.3	29.1	11.8
Sep 81	36.	32.4	20.8	92.5	18.5	12.1
Oct 81	41.1	34.1	49.5	93.8	33.3	12.8
Nov 81	70.	61.4	68.5	95.9	51.7	13.
Dec 81	64.6	62.	53.1	93.8	54.8	13.5
Jan 82	52.2	59.2	50.	96.7	66.3	13.5
Feb 82	50.5	41.1	63.	97.	67.7	14.1
Mar 82	54.9	63.8	74.1	97.1	67.	14.2
Apr 82	58.5	46.4	55.	90.4	37.3	14.7
May 82	70.	43.3	62.5	97.6	56.2	14.3
Jun 82	63.9	49.1	46.	91.2	28.9	14.4
Jul 82	52.5	37.1	23.1	96.2	28.9	14.5
Aug 82	50.	42.6	5.7	92.2	15.4	15.2
Sep 82	39.	37.5	2.3	85.8	17.7	15.3
Oct 82	61.7	55.8	29.8	91.8	32.3	15.9
Nov 82	48.3	55.9	37.8	93.	39.1	18.
Dec 82	58.6	62.5	29.4	97.3	47.9	17.8
Jan 83	52.5	37.	39.	87.9	30.3	17.6
Feb 83	68.1	69.8	44.4	99.5	45.1	17.8
Mar 83	74.7	55.8	37.9	96.8	41.8	16.6
Apr 83	68.1	50.	54.5	98.4	46.7	17.
May 83	65.9	75.	47.	91.9	41.9	17.6
Jun 83	66.3	55.8	40.3	92.	36.1	15.7
Jul 83	53.9	53.2	22.8	95.2	37.	15.7
Aug 83	67.4	70.2	53.8	100.	51.7	15.9
Sep 83	82.8	68.8	46.5	98.1	62.3	15.
Oct 83	60.7	64.6	41.5	95.9	42.6	14.8
Nov 83	55.6	69.8	41.2	95.6	56.5	14.8
Dec 83	66.	70.8	52.6	95.5	48.	13.7
Jan 84	68.8	72.	52.7	96.2	52.	14.2
Feb 84	59.1	66.7	48.9	94.2	55.6	14.0
Mar 84	51.8	75.	62.5	95.9	61.8	13.8

APPENDIX B
THE BOX-JENKINS METHOD

A. OVERVIEW

The Box-Jenkins procedure can be used to fit and forecast time series data by means of a general class of statistical models. An observation at a given point in time is modeled as a function of its past values and/or current and past values of the random errors, both at seasonal and non-seasonal lags. Box-Jenkins methodology will model a variable with observations equally spaced in time and no missing values. Sometimes it may be necessarry, before modeling the series, to transform the data by taking the log function, square root, power of the series or to difference the series on a seasonal or non-seasonal basis.

The modeling of time series data is usually done in three steps. First, identify a tentative model for the series. Second, estimate the parameters[3] and examine the diagnostic plots and statistics. Third, if the model is deemed acceptable, utilize the procedure for forecasting. If the model is inadequate, return to step one and evaluate the time series for more appropriate models until an acceptable one is found. Fig. B.1 illustrates the steps required in Box-Jenkins analysis.

The advantage gained in using Box-Jenkins analysis is that it allows the data to speak for itself since it is a univariate procedure and therefore does not allow for explanatory variables. The underlying more restrictive assumption

---

[3]For most software packages applied to Box-Jenkins modeling, the estimation of parameters is done automatically in the program leaving the researcher free to concentrate on analysis of the resulting statistics.

throughout all Box-Jenkins procedures is that the time series will eventually repeat itself [Ref. 7], or that there is some pattern underlying the data.

B. WHAT IS A TIME SERIES

A time series is a collection of observations generated sequentially over time at specific intervals such as hours, days, weeks, months or years. In addition, a certain dependence is supposed from one period to the next. It is this interdependence that is of value when trying to forecast future activity for a time series. Examples of time series abound in fields ranging from business to physics and are applied to analyze monthly sales for a company, quarterly yields on fiduciary notes or the chemical yield of a certain substance in a controlled procedure. A time series can also be used to analyze observations that are either discrete or continuous; by way of example, a discrete time series would be the closing stock price of a company and a continuous time series would be the temperature at the weather center. In summary, a discrete time series is one where observations exist at a point in time and a continuous time series has potential observations at all points in time. For purposes of this discussion, only discrete time series are be addressed.

C. STATIONARITY

In pursuing a time series model, the first assumption to be made in the analysis is that the data set is stationary. By this it is meant that the observations oscillate around a constant mean that shows no growth over time. Deviations about this mean are temporary and in the long run display equilibrium about the mean [Ref. 4].

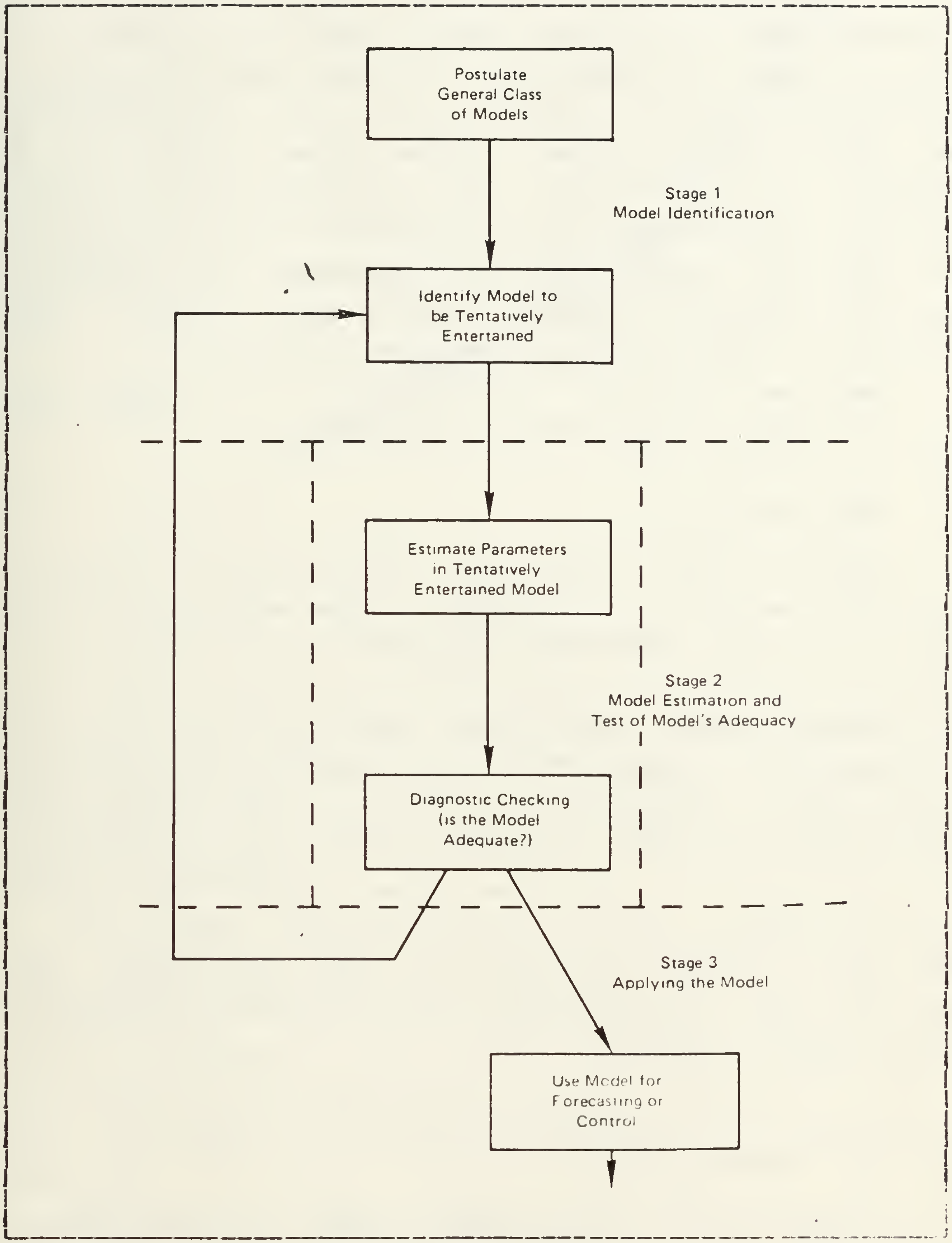

Figure B.1 THE BOX-JENKINS METHODOLOGY.

A further measure of stationarity can be gained from the autocorrelation of the time series, that is the correlation between successive observations from the same data set. An observation at time t, denoted by Zt, when correlated with an observation Zt+1 from the same data set is said to produce an autocorrelation. The autocorrelation is measured by pk and provides important information about the nature of the data set. A value close to +1 indicates a high degree of positive correlation between observations, while a value close to -1 indicates a high negative correlation.

Most time series are not stationary and require some type of transformation prior to analysis.

D. TIME SERIES PLOT

The first step in determining whether a time series is stationary or not is to construct a time series plot of the data which plots observations against time in an attempt to visually determine any obvious patterns in the data. Fig. B.2 illustrates the United States gross national product for the years 1947 through 1970 on a quarterly basis. This plot shows a clear upward trend in the data which indicates the data is not stationary, further within each year there are apparently recurring patterns for each quarter that repeat on annual basis. Finally, as time passes, the variance in GNP tends to become larger and more volatile. Clearly, this data set must be transformed prior to further analysis by the Box-Jenkins method.

E. DATA TRANSFORMATION

To continue with the example of GNP, there are several possible transformations that can be used to induce stationarity. The first step is to induce a constant variance in

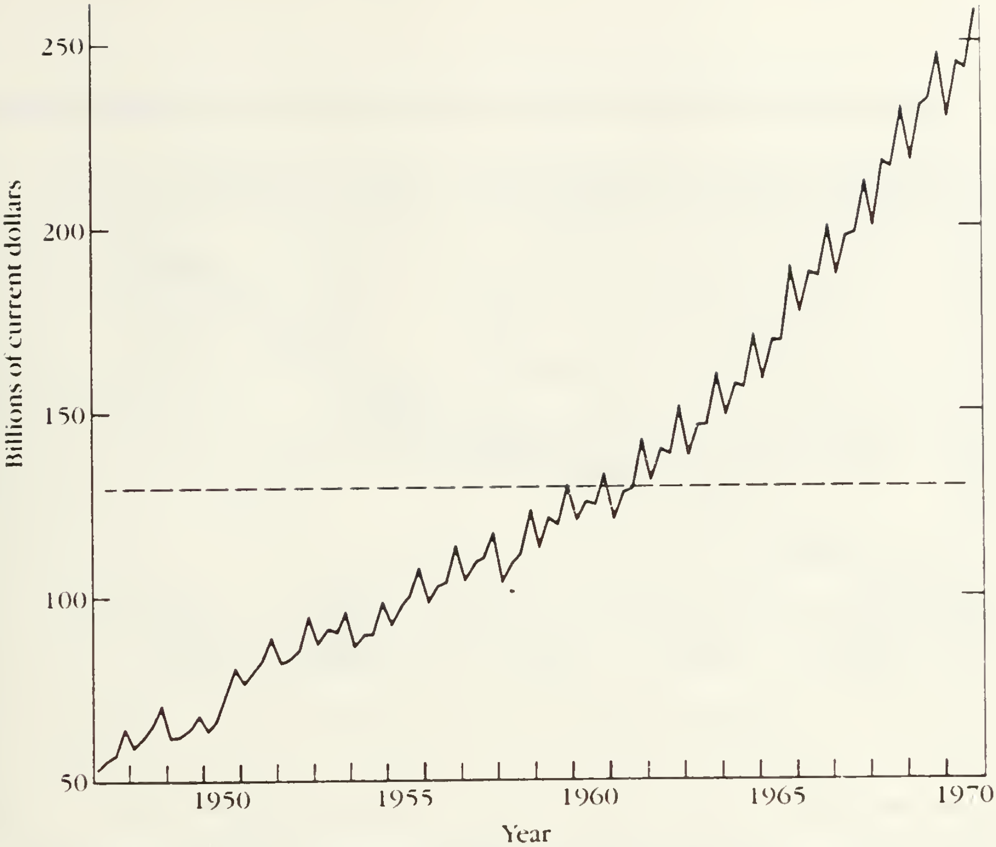

Figure B.2 UNITED STATES GNP 1947 - 1970.

the data, this can be accomplished either through a logarithmic or a square root transformation. Fig. B.3 illustrates the results of a square root transform of the data. The trend is still clearly present but the variance has been smoothed considerably.

Once variance has been stabilized, the next step is to remove the trend. There are several sophisticated regression techniques available to accomplish this however, the method of differencing will be the only one addressed here. For a more detailed discussion of these alternative techniques, the reader is directed to Makridakis and Wheelright [Ref. 4].

The method of differencing a time series consists of subtracting the values of the time series from each other in a specified order. By way of example, consider the following

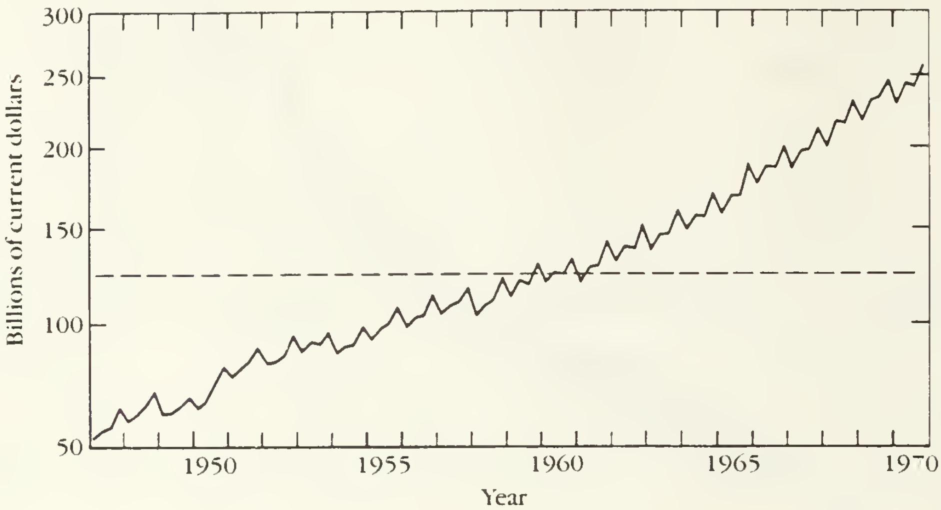

Figure B.3 SQUARE ROOT TRANSFORMATION OF U.S. GNP 1947 - 1970.

data set; 1,3,5,7,9,11. When plotted, this set has an obvious linear trend, a first order difference involving subtraction of the first observation from the second, the second from the third and so on results in the transformed series *,2,2,2,2,2. By taking the first difference, the trend disappears and yields a stationary data set. The '*' indicates that whenever a data set is differenced, one observation is lost for each difference operator. Due to random fluctuations in real data, such clear cut results as those illustrated should not be anticipated however, for the majority of data sets differencing will induce a sufficient amount of stationarity to proceed with further analysis.

F. AUTOCORRELATION AND PARTIAL AUTOCORRELATION

A useful tool in model estimation is the autocorrelation function (ACF), which can be defined as the association or mutual dependence between values of the same variable but at different time periods. These ACF coefficients provide valuable information about a data set and any pattern that may be present. If, for example, a high positive coefficient appeared every twelve months, a seasonal trend may be considered to exist.

The partial autocorrelation function, (PACF), is another and complimentary measure to be applied along with the ACF to aid in determining model type. PACF's are analogous to ACF's in that they indicate the relationship of the values of a time series to various time lagged values of the same series. They differ from ACF's, however, in that they are computed for each time lag after removing the effect of all other time lags. In essence, they show the relative strength of the relationship that exists for varying time lags.

When the ACF and the PACF are analyzed together, they provide a very powerful tool for initial model selection. Fig. B.4 through Fig. B.6 summarize the general shapes associated with the different types of models.

G. THE AUTOREGRESSIVE MODEL

A time series is said to be governed by an autoregressive, (AR), if the current value of the time series can be expressed as a linear function of the previous value or values plus some error term or random shock value [Ref. 8]. The assumptions made here are that the data set is stationary and the error terms are normally and independently distributed with a mean of zero and constant variance. A check on the adequacy of the model is to construct an ACF for the residuals of the model and determine if they

are random in nature. Mathematically, an AR (1) model is of the form:

$$Z_t = \phi Z_{t-1} + a_t \qquad \text{(eqn B.1)}$$

where ϕ is equal to the autoregressive coefficient and a is the random error or shock term.

H. THE MOVING AVERAGE MODEL

A time series is said to be governed by a moving average process if the current value of the time series can be expressed as a linear function of the current error term and previous error term(s). The same restrictions apply to the error terms of am MA model as applied to the AR model. Mathematically, this function can be expressed as;

$$Z_t = a_{t-1} - \theta a_t \qquad \text{(eqn B.2)}$$

where θ is the moving average coefficient and a is again the error term.

I. THE MIXED AUTOREGRESSIVE MOVING AVERAGE MODEL

The mixed autoregressive moving average model contains elements of both the AR and the MA procedures and expresses the relationship of a current observation as a linear function of both past values and past errors of the variable. As with the AR and MA models, the residuals of the model are evaluated for adequacy by utilizing the ACF function. The equation for the ARMA model is:

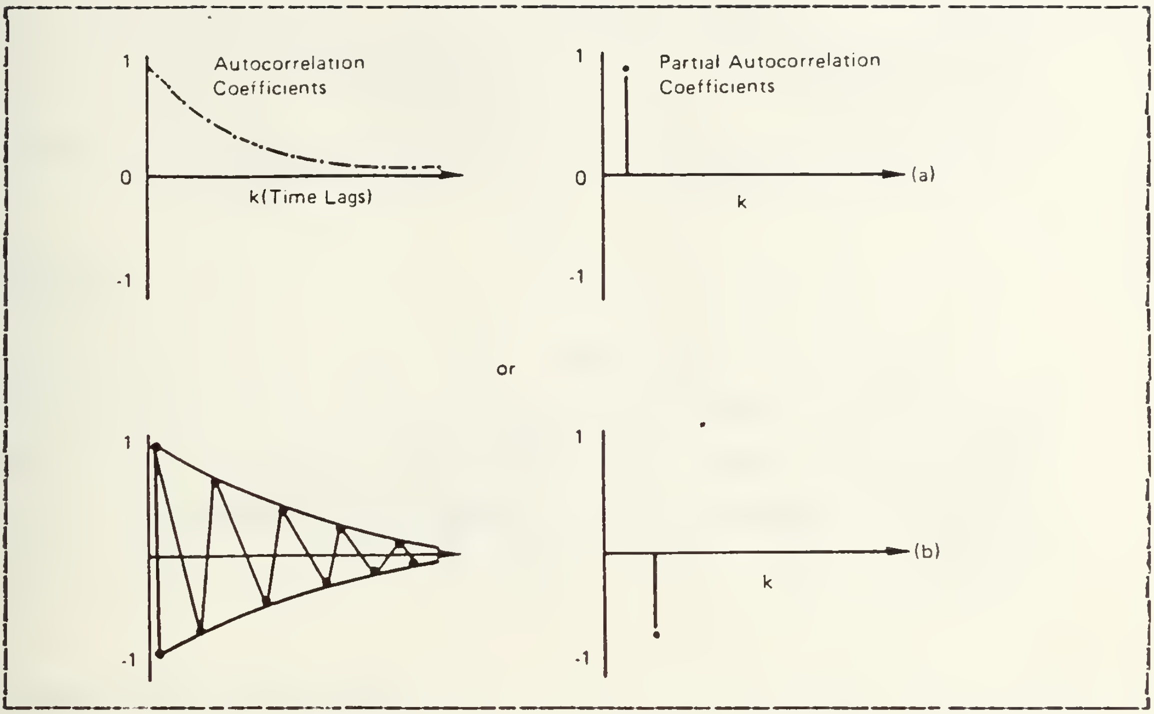

Figure B.4 TYPICAL FORM OF AR1 MODEL ACF AND PACF.

$$Z_t = \phi_1 Z_{t-1} + a_t - \theta a_{t-1} \qquad \text{(eqn B.3)}$$

J. EVALUATING THE MODEL

Once a model has been selected, there are several ways to check the results for adequacy. For purposes of this discussion, the following checks will be addressed:

1. ACF of residuals
2. Minimum sum of squares
3. t-ratio

There are several other checks for adequacy that are available to the user of Box-Jenkins methodology, for a more comprehensive explanation the reader is directed to Vandaele [Ref. 8].

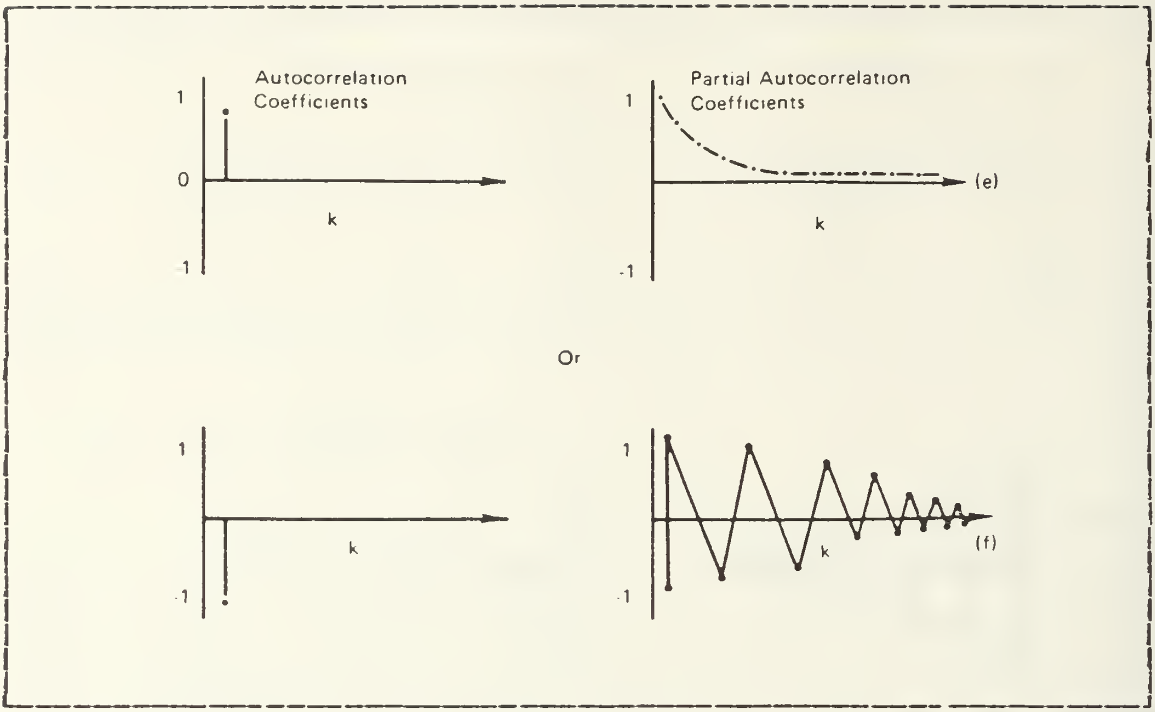

Figure B.5 TYPICAL FORM OF MA1 MODEL ACF AND PACF.

1. ACF of The Residuals

As mentioned throughout this discussion, the ACF for a model should be random about the mean zero with constant variance and a magnitude less than two standard errors.

2. Minimum Sum of Squares

Determination of this measure can only be achieved by comparison with other potential models. In some cases, several models may produce insignificantly different sums of squares which will test the user's judgment and application of the other measures of adequacy.

3. t-Ratio

In time series analysis this is computed by dividing the estimate of the parameter for the model by the standard

deviation for the series. The rules as applied to regression analysis still hold in that the value should be greater than +/-2.0 in order to indicate that the coefficient is significantly different from zero.

K. PARSIMONY

In the event that more than one model is capable of satisfying the acceptance criteria described above, the principle of parsimony will then apply. This states that when faced with several sufficient model types, select the the lowest order model available that satisfies the criteria.

L. TRANSFER FUNCTION MODELS

Also known as multivariate autoregressive integrated moving average, (MARIMA), or leading indicator models. This involves selection of an appropriate univariate model for what is to be the independent variable and applying it to the dependent variable. The application of the model will result in two sets of residuals which when cross correlated at different time lags will yield the cross correlation function, (CCF). This differs slightly from the ACF discussed earlier in that we now expect to find a relationship of significant magnitude at various points in the comparison. This positive correlation is an indication that there is a significant relationship between the independent and the dependent variable at certain lags or leads in time.

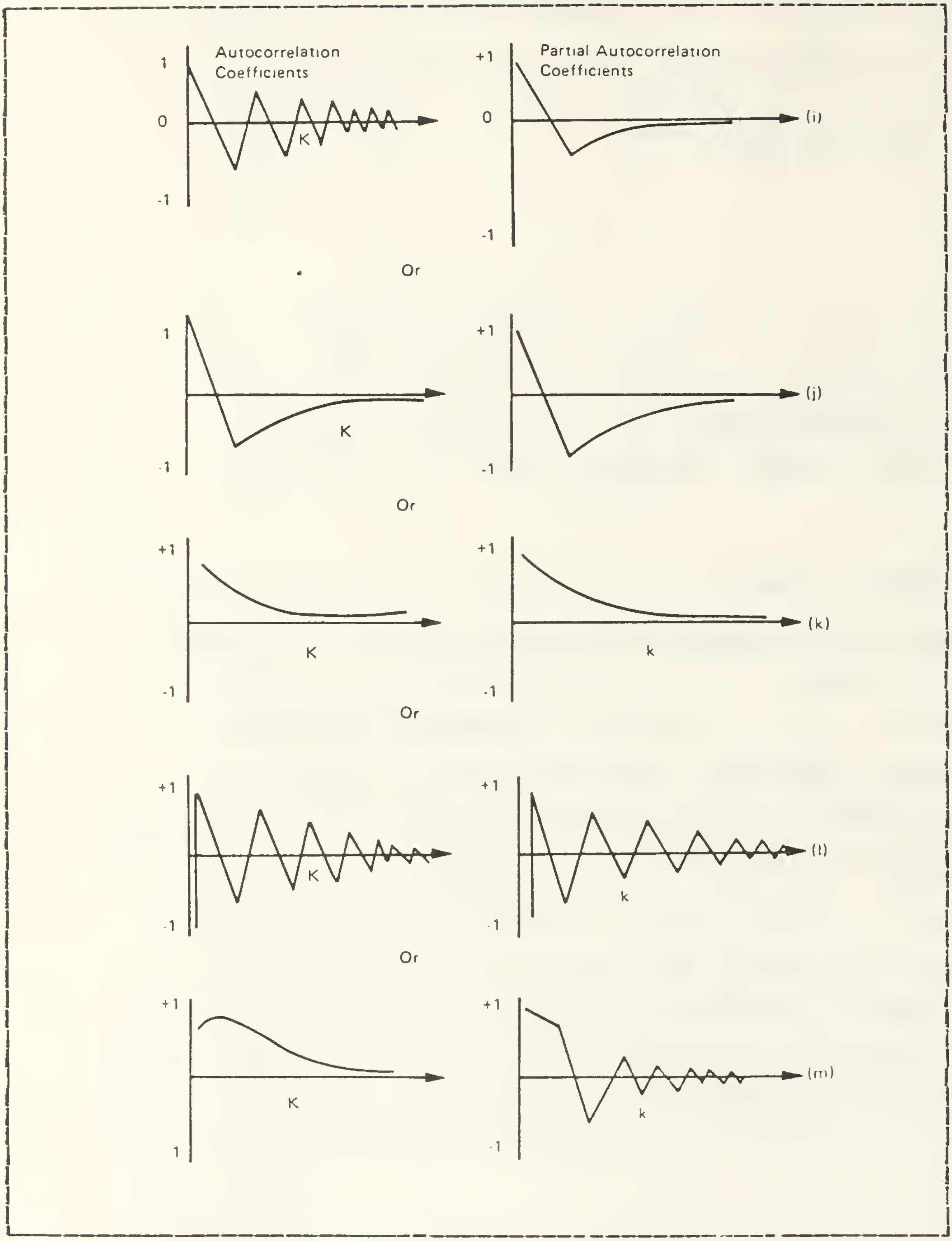

Figure B.6 TYPICAL FORM OF ARMA 1,1 MODEL ACF AND PACF.

APPENDIX C

SUMMARY OF BOX-JENKINS MODELS USED FOR ANALYSIS

A. MODELS USED IN UNIVARIATE ANALYSIS

1. Yeoman

Model Type - ARIMA (1,0,0)

Model Equation

$$Z = .448z_{t-1} + a_t \qquad \text{(eqn C.1)}$$

T-ratio - 2.64

Model Residuals - Random

2. Storekeepers

Model Type - ARIMA (0,1,1)

Model Equation

$$Z = -.814a_{t-1} + a_t \qquad \text{(eqn C.2)}$$

T-ratio - 7.24

Model Residuals - Random

3. <u>Operations Specialists</u>

Model Type - ARIMA (1,0,0)

Model Equation

$$Z = .499Z_{\tau-1} + a_\tau \qquad \text{(eqn C.3)}$$

T-ratio - 3.39

Model Residuals - Random

4. <u>Electronics Technicians</u>

Model Type - ARIMA (1,0,0)

Model Equation

$$Z = .588Z_{\tau-1} + a_\tau \qquad \text{(eqn C.4)}$$

T-ratio - 4.18

Model Residuals - Random

5. Boiler Technicians

Model Type - ARIMA (1,0,0)

Model Equation

$$Z = .541Z_{t-1} + a_t \qquad \text{(eqn C.5)}$$

T-ratio - 3.63

Model Residuals - Random

B. BOX-JENKINS MODELS OF REGRESSION RESIDUALS

1. Yeoman

Model Type - ARIMA (0,0,1)

Model Equation

$$Z = -.376a_{t-1} + a_t \qquad \text{(eqn C.6)}$$

T-ratio - -2.19

Model Residuals - Random

2. Storekeepers

Model Type - ARIMA (1,1,0)

Model Equation

$$Z = .498Z_{t-1} + a_t \qquad \text{(eqn C.7)}$$

T-ratio - -3.00

Model Residuals - Random

3. Operations Specialists

Model Type - ARIMA (1,1,1)

Model Equation

$$Z = .604Z_{t-1} + a_t - .956a_{t-1} \qquad \text{(eqn C.8)}$$

T-ratio - AR1 - 3.28
MA1 - 9.75

Model Residuals - Random

4. <u>Electronics Technicians</u>

Model Type - ARIMA (1,0,1)

Model Equation

$$Z = .854Z_{t-1} + a_t - .529a_{t-1} \qquad \text{(eqn C.9)}$$

T-ratio - AR1 - 5.21
MA1 - 2.04

Model Residuals - Random

5. <u>Boiler Technicians</u>

Model Type - ARIMA (1,0,0)

Model Equation

$$Z = .590Z_{t-1} + a_t \qquad \text{(eqn C.10)}$$

T-ratio - 3.70

Model Residuals - Random

LIST OF REFERENCES

1. Bepko, J.J., III, Econometric Models of U.S.Navy Career Petty Officer Retention Masters Thesis, Naval Postgraduate School, Monterey, 1981.

2. Darling, P., An Analysis and Forecast of the Supply of First Term Enlistees to the United States Marine Corps, Masters Thesis, Naval Postgraduate School, Monterey, 1979.

3. Thomas, G.W. and Liao, S.S., A Personnel Planning Model: Estimating the Turnover Among Experienced Military Employees, Naval Postgraduate School Research Foundation, May, 1984.

4. Makridakis, S., and Wheelwright, S., Interactive Forecasting: Univariate and Multivariate Methods, 2d ed., Holden-Day, 1978.

5. Cowin, M., O'Connor, F.E., Sage, L.C. and Johnson, C.R., The Effects of Local Economic Conditions on Navy Enlistment Information Spectrum Inc., March 1980

6. Pindyck, R.S. and Rubinfeld, D.L., Econometric Models and Economic Forecasts, 2d ed., McGraw Hill, 1981

7. SPSS Inc., SPSSx Users Guide McGraw Hill, 1983

8. Vandaele, W., Applied Time Series and Box-Jenkins Models, Academic Press, inc., 1983.

9. Ryan, T., Joiner, B., and Ryan, B., Minitab Reference Manual, Penn State University, 1981.

INITIAL DISTRIBUTION LIST

	No. Copies
1. Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2. Library, Code 0142 Naval Postgraduate School Monterey, California 93943	2
3. Chief of Naval Operations (OP-01,11,12,13) Department of the Navy Washington, D.C. 20350	4
4. Department Chairman, Code 54 Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	1
5. Professor G. Thomas, Code 54TC Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	2
6. Professor W. Mc Garvey, Code 54Ms Department of Administrative Science Naval Postgraduate School Monterey, California 93943	1
7. LCDR K.J. Sherry USS Jouett (CG-29) FPO San Francisco 96669	5

210589

Thesis
S4497 Sherry
c.1
A combined regression and Box-Jenkins prediction model for reenlistment in selected Navy ratings.